

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:21:36 ; Search time 90 Seconds
(without alignments)
2054.627 Million cell updates/sec

Title: US-10-016-768A-10
Perfect score: 6030
Sequence: 1 MHSSYSISLSEKVAEECWGR.....ILHEKLAQIKAEVDAQDQL 1165

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_19Jun03.*
2: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6030	100.0	1165	23	AAE24372
2	5882	97.5	1140	22	ABR71145
3	4839	80.2	1221	23	AAE24373
4	314.5	5.2	1937	22	ABBS58985
5	314.5	5.2	2151	22	ABR60086
6	311.5	5.0	2112	22	ABR60403
7	299.5	5.0	2175	22	ABR65698
8	296.5	4.9	1161	22	ABR70667
9	294	4.9	2441	22	ABR62231

10	293	4.9	1436	22	ABR61959	Drosophila melanog
11	287	4.8	1369	22	ABR60040	Drosophila melanog
12	285.5	4.7	1409	22	ABR61328	Drosophila melanog
13	285	4.7	1703	22	ABR66223	Drosophila melanog
14	283	4.7	1416	22	ABR65499	Drosophila melanog
15	282.5	4.7	1741	22	ABR60091	Drosophila melanog
16	281.5	4.7	1198	22	ABR60371	Drosophila melanog
17	281	4.7	2280	22	ABR61650	Drosophila melanog
18	279.5	4.6	1186	22	ABR63516	Drosophila melanog
19	279	4.6	2110	22	ABR58077	Drosophila melanog
20	279	4.6	2150	23	AAO22566	Wooden leg (WOL) g
21	278.5	4.6	1665	22	ABR64010	Drosophila melanog
22	278	4.6	53	23	AAE24370	Fruit fly E93 prog
23	276	4.6	1078	22	ABR58620	Drosophila melanog
24	274.5	4.6	3502	22	ABR58382	Drosophila melanog
25	274	4.5	2344	22	AAU37120	Staphylococcus aur
26	273.5	4.5	1164	22	ABR57802	Drosophila melanog
27	272.5	4.5	1246	22	ABR61547	Drosophila melanog
28	272.5	4.5	1669	22	ABR64003	Drosophila melanog
29	271	4.5	2309	22	ABR66232	Drosophila melanog
30	270.5	4.5	1217	24	ABR60377	Bombyx mori serici
31	270	4.5	1467	23	ABR97605	Novel human protei
32	268.5	4.5	1368	22	ABR60262	Drosophila melanog
33	267.5	4.4	1920	22	ABR64441	Drosophila melanog
34	266	4.4	1028	22	ABR62708	Drosophila melanog
35	266	4.4	2283	22	ABR66876	Staphylococcus epi
36	263.5	4.4	2261	24	ABJ18914	Pathogen specific
37	262	4.3	1911	22	ABR61038	Drosophila melanog
38	261.5	4.3	1326	22	ABR61156	Drosophila melanog
39	261	4.3	2451	22	ABR71574	Drosophila melanog
40	259.5	4.3	1545	22	ABR65577	Drosophila melanog
41	257	4.3	738	22	ABR58805	Drosophila melanog
42	257	4.3	1001	22	ABR60527	Drosophila melanog
43	257	4.3	1805	22	ABR65262	Drosophila melanog
44	256.5	4.3	1059	22	ABR65792	Drosophila melanog
45	256.5	4.3	1077	22	ABR61539	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAE24372	
ID	AAE24372 standard; Protein: 1165 AA.
AC	AAE24372;
XX	
DT	
DE	04-OCT-2002 (first entry)
XX	
XX	Fruit fly E93 programmed cell death modulating protein #1.
KW	Fruit fly; programmed cell death modulating protein; adenocarcinoma;
KW	cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW	neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
KW	Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
KW	aplastic anaemia; ischemic injury; myocardial infarction; stroke;
KW	reperfusion injury; toxin-induced disease; genetic immunodeficiency;
KW	vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;
KW	vaccine; noctropic; vasotropic; immunostimulant; cerebroprotective;
KW	cardiant; cancer; E93 protein.
OS	
XX	Drosophila melanogaster.
PN	WO200234882-A2.
XX	
PD	02-MAY-2002.
XX	
PF	29-OCT-2001; 2001WO-US48053.
XX	
PR	27-OCT-2000; 2000US-243865P.
XX	
PA	(UYWA-) UNIV MARYLAND BIOTECHNOLOGY INST.
XX	

PI Baehrecke EH;
 XX WPI: 2002-479717/51.
 DR N-PSDB; AADJ39237.
 XX Novel programmed cell death modulating proteins, useful for treating or
 PT preventing disorders associated with abnormal cell proliferation and
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial
 PT infarction -
 XX
 PS Claim 9; Page 65-71; 88pp; English.
 CC The present invention relates to novel programmed cell death modulating
 CC proteins and polynucleotides encoding such proteins. Sequences of the
 CC invention are useful to screen potential cellular apoptosis inhibiting
 CC compounds to determine their use as therapeutic agents for treatment of
 CC diseases associated with increased programmed cell death. They are also
 CC useful for treating or preventing disorders associated with decrease in
 CC apoptosis. Programmed cell death modulating sequences are useful for
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the
 CC invention are useful for treating disorders associated with increase
 CC in cell death or apoptosis such as acquired immunodeficiency syndrome
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced
 CC diseases and other infectious or genetic immunodeficiencies. Sequences
 CC of the invention are used as vaccines and in gene therapy. The present
 CC sequence is fruit fly E93 programmed cell death modulating protein.
 XX
 SO Sequence 1165 AA:
 Query Match 100.0%; Score 6030; DB 23; Length 1165;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHISYEELERVAEECGRROWKHODKLTCSHINIEEQPIAAGSEDEPSOYNHSSK 60
 DB 1 MHISYEELERVAEECGRROWKHODKLTCSHINIEEQPIAAGSEDEPSOYNHSSK 60
 QY 61 EISQSNPNHCKTENHRLTEQHNHSGQLLEEDSENNQTSRTPPTSPSPPEPI 120
 DB 61 EISQSNPNHCKTENHRLTEQHNHSGQLLEEDSENNQTSRTPPTSPSPPEPI 120
 QY 121 DMRPSAKCNFCVNGRLTVNAOGKLVASATATSSSTSHIHQSDSNSASLPHHI 180
 DB 121 DMRPSAKCNFCVNGRLTVNAOGKLVASATATSSSTSHIHQSDSNSASLPHHI 180
 QY 181 SSSSSNNNSGNNRARIIAASARATPAATPANSLEYLTLTORAAKMTSMDSMAAOLA 240
 DB 181 SSSSSNNNSGNNRARIIAASARATPAATPANSLEYLTLTORAAKMTSMDSMAAOLA 240
 QY 241 QFSLLADENLINSLASQQQQQQQQQIAGAVPTTSEVSAAIISPAKXTPSPVADPDL 300
 DB 241 QFSLLADENLINSLASQQQQQQQQQIAGAVPTTSEVSAAIISPAKXTPSPVADPDL 300
 QY 301 SSKSPSNSISGDVYKSVACATPTPSGRRAVSEEDLSALODVVANKLDARKSASQHHQ 360
 DB 301 SSKSPSNSISGDVYKSVACATPTPSGRRAVSEEDLSALODVVANKLDARKSASQHHQ 360
 QY 361 RSIIDNRLFKMKHHDOEDHGDDESDNDAAEAVDSNASTPVYPAFAAQAQRKLSHL 420
 DB 361 RSIIDNRLFKMKHHDOEDHGDDESDNDAAEAVDSNASTPVYPAFAAQAQRKLSHL 420
 QY 421 SEHNGSDLDGEDVDGRSPKGRHAPACGNASANOAGAPASIPLDANVLLHTLMTAAGIGAMPK 480
 DB 421 SEHNGSDLDGEDVDGRSPKGRHAPACGNASANOAGAPASIPLDANVLLHTLMTAAGIGAMPK 480
 QY 481 LDETQTVGDFIKGLIVANSGGIMNEGILNLLSASGENSNGNASLLQQQOHQOHQOHQ 540
 DB 481 LDETQTVGDFIKGLIVANSGGIMNEGILNLLSASGENSNGNASLLQQQOHQOHQOHQ 540

QY 541 QQQQOHHVAAVRRHLPKSETPETNSSLPDNDASEDPILKIPSFVSGPASSSSLSPGGLV 600
 DB 541 QQQQOHHVAAVRRHLPKSETPETNSSLPDNDASEDPILKIPSFVSGPASSSSLSPGGLV 600
 QY 601 GGHHPPLNNNSLSISNNNSHSSNSHRNGSNRSPHSASPMIAAAVAGCYAGNSLLTSS 660
 DB 601 GGHHPPLNNNSLSISNNNSHSSNSHRNGSNRSPHSASPMIAAAVAGCYAGNSLLTSS 660
 QY 661 SSSIQKMASTIQOINQSQESLERNGVSDCSNNGSSSLGKPKPSISVAKIIGTD 720
 DB 661 SSSIQKMASTIQOINQSQESLERNGVSDCSNNGSSSLGKPKPSISVAKIIGTD 720
 QY 721 TSRCGAPNLLSQOHHSAHLTHQOQQOOLSAQALGKTPPKGKRRNDRDRLVEAVK 780
 DB 721 TSRCGAPNLLSQOHHSAHLTHQOQQOOLSAQALGKTPPKGKRRNDRDRLVEAVK 780
 QY 781 AVORGEMSVHRAGSYGVPHSTLEYKVERHLMRPRKREPPODLVGLTGPAKQLDK 840
 DB 781 AVORGEMSVHRAGSYGVPHSTLEYKVERHLMRPRKREPPODLVGLTGPAKQLDK 840
 QY 841 LKAGPHGSKLSNALKNNQNNQAAAAAATPNGIKLPLEFAGPOLASFOPMF 900
 DB 841 LKAGPHGSKLSNALKNNQNNQAAAAAATPNGIKLPLEFAGPOLASFOPMF 900
 QY 901 WPTNATNAYGLDFNRITTEAMRNPOASNHGKMSAODMENVYDGIIRKTLQASENGS 960
 DB 901 WPTNATNAYGLDFNRITTEAMRNPOASNHGKMSAODMENVYDGIIRKTLQASENGS 960
 QY 961 AAGNGSNGSNGHGHGHGHALLDQLVYKTPPLPTNRHNDNYATCSSASGESYKRS 1020
 DB 961 AAGNGSNGSNGHGHGHGHALLDQLVYKTPPLPTNRHNDNYATCSSASGESYKRS 1020
 QY 1021 GSPWGNVADIKREKLSADSGSSSDEHSASHINNNSDLAHNKKSGGGGGGNGCOTNG 1080
 DB 1021 GSPWGNVADIKREKLSADSGSSSDEHSASHINNNSDLAHNKKSGGGGGGNGCOTNG 1080
 QY 1081 NGRSSRMTSRDSDSTDPASFPKSGENGGOONHKKMDLNGSSSSSHICESEPAATGHHSPG 1140
 DB 1081 NGRSSRMTSRDSDSTDPASFPKSGENGGOONHKKMDLNGSSSSSHICESEPAATGHHSPG 1140
 QY 1141 HTTSTLHEKLAQIKAEQVDQADQL 1165
 DB 1141 HTTSTLHEKLAQIKAEQVDQADQL 1165
 RESULT 2
 ABB71145
 ID ABB71145 standard; Protein; 1140 AA.
 AC ABB71145;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 40227.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 OS
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 RR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li FMD, Myers EW,
 XX

DR WPI: 2001-65660/75.
 DR N-PSDB; ABL15248.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclousure; SEQ ID NO 40227; 21np + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB2072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 1140 AA;
 XX
 SQ
 Query Match 97.5%; Score 5882; DB 22; Length 1140;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1135; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Oy 18 MGRQWKHYQDKLTCTSHNIEEQPIAAGSEDEQVNHSSKEISQSNPNHCKTENRRL 77
 Db 1 MGRQWKHYQDKLTCTSHNIEEQPIAAGSEDEQVNHSSKEISQSNPNHCKTENRRL 60
 Oy 78 EQHNGSOLLEBEDENNQTSHTSRTPTPGATSPSPPEPIDMRPSAKCFVNGRL 137
 Db 61 EQHNGSOLLEBEDENNQTSHTSRTPTPGATSPSPPEPIDMRPSAKCFVNGRL 120
 Oy 138 TVNAGKLVASAAATATSSSTNSHIIHQHSDNSASLPHIISSSSSNNNSGNRAH 197
 Db 121 TVNAGKLVASAAATATSSSTNSHIIHQHSDNSASLPHIISSSSSNNNSGNRAH 180
 Oy 198 IAAASARTPAATPANSLELYKLTORAAKMTSDMSAAQLAQSLLADFLINLSAQ 257
 Db 181 IAAASARTPAATPANSLELYKLTORAAKMTSDMSAAQLAQSLLADFLINLSAQ 240
 Oy 258 QQQQQQQQIASAIVPTTSEVSAATSPALKDTPSPSVAPLDLSKSPNSISGDVSV 317
 Db 241 QQQQQQQQIASAIVPTTSEVSAATSPALKDTPSPSVAPLDLSKSPNSISGDVSV 300
 Oy 318 RACATPTSGRRAYSEEDLSRALQDVANKLDARKSASQHEORSILDNRLFQKHGHOE 377
 Db 301 RACATPTSGRRAYSEEDLSRALQDVANKLDARKSASQHEORSILDNRLFQKHGHOE 360
 Oy 378 ODHGDDELSDNDAAEAVDSNAATPVYPAEFARQKLKLSLHSHNSDLDGEDVDRSP 437
 Db 361 ODHGDDELSDNDAAEAVDSNAATPVYPAEFARQKLKLSLHSHNSDLDGEDVDRSP 420
 Oy 438 KMGRRPACGNASANOAGAPASIPLDANVLLHTMLAAGIGAMPKLDCTQVGFIKGLVA 497
 Db 421 KMGRRPACGNASANOAGAPASIPLDANVLLHTMLAAGIGAMPKLDCTQVGFIKGLVA 480
 Oy 498 NSGIMNEGLNLTLASQENSNGNASLLLQQQOHOHQHQQOQQOQQOQVAAVYHRLLPK 557
 Db 481 NSGIMNEGLNLTLASQENSNGNASLLLQQQOHOHQHQQOQQOQQOQVAAVYHRLLPK 540
 Oy 558 SETPETNSLDPNDA SEDPIKIPSFKVSGPASSSLSPGGLVGHHHPPLNNNSLSISN 617
 Db 541 SETPETNSLDPNDA SEDPIKIPSFKVSGPASSSLSPGGLVGHHHPPLNNNSLSISN 600
 Oy 618 NSNHSNHRNGSNNSPHSASPMLAAVAAQGSYSGNLSLTSSSSSIKMAASNIQORIN 677
 Db 601 NSNHSNHRNGSNNSPHSASPMLAAVAAQGSYSGNLSLTSSSSSIKMAASNIQORIN 660
 Oy 678 EQSGGESLRNGNVSDCSSNNNGSSSLGYKKPSISVAKIIGGDTISRFGASPILLSQOHS 737
 XX

Db 661 EQSGGESLRNGNVSDCSSNNNGSSSLGYKKPSISVAKIIGGDTISRFGASPILLSQOHS 720
 Oy 738 AHHLTHQOOQOOLSAOELGKGTBPKRGKYNNYRDSIYEAVKAVRGEMSVHRAGSYYG 797
 Db 721 AHHLTHQOOQOOLSAOELGKGTBPKRGKYNNYRDSIYEAVKAVRGEMSVHRAGSYYG 780
 Oy 798 VPHSTLEYKVERHLMRPRKREPQPDVLVGLTGPANKLQDLKAKGPHGSKLSNALKN 857
 Db 781 VPHSTLEYKVERHLMRPRKREPQPDVLVGLTGPANKLQDLKAKGPHGSKLSNALKN 840
 Oy 858 QNNQAAAAAATAATNGCLKPLFEAGPOLSFQPMFVFNQTNATNAYGIDFNRI 917
 Db 841 QNNQAAAAAATAATNGCLKPLFEAGPOLSFQPMFVFNQTNATNAYGIDFNRI 900
 Oy 918 TEARNPNPASHHGMASAOQVENVYDGIIRKTLQASSEGNSAAGNSNSNGHGHG 977
 Db 901 TEARNPNPASHHGMASAOQVENVYDGIIRKTLQASSEGNSAAGNSNSNGHGHG 960
 Oy 978 HGHGHALLDQLLVKKTPLPTNHRNDYATCSSASGSESVRSRSGPMGNVADIKERLSA 1037
 Db 961 HGHGHALLDQLLVKKTPLPTNHRNDYATCSSASGSESVRSRSGPMGNVADIKERLSA 1020
 Oy 1038 DSGSSDEHSHASHINNNSDLAHNKNKSGGGGGGNGGTNGRSGSRMTSRDDSETDA 1097
 Db 1021 DSGSSDEHSHASHINNNSDLAHNKNKSGGGGGGNGGTNGRSGSRMTSRDDSETDA 1080
 Oy 1098 SSFSGENGGQONKMDLNGSSSSSHIKESAAATHSHSPGHTTSLIHEKLAQICAE 1157
 Db 1081 SLSKSGESGGQONKMDLNGSSSSSHIKESAAATHSHSPGHTTSLIHEKLAQICAK 1140
 XX
 XX RESULT 3
 XX AAE24373
 XX ID AAE24373 standard; Protein; 1221 AA.
 XX AC AAE24373;
 XX DT 04-OCT-2002 (first entry)
 XX
 XX Fruit fly E93 programmed cell death modulating protein #2.
 XX
 XX Fruit fly; programmed cell death modulating protein; adenocarcinoma;
 XX cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
 XX neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;
 XX Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;
 XX aplastic anaemia; ischaemic injury; myocardial infarction; stroke;
 XX reperfusion injury; toxin-induced disease; genetic immunodeficiency;
 XX vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;
 XX myeloma; nocotropic; vasotropic; immunostimulant; cerebroprotective;
 XX cardiac; cancer; E93 protein.
 XX
 XX Drosophila melanogaster.
 XX
 XX WO200224882-A2.
 XX
 XX 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-US48053.
 XX
 XX 27-OCT-2000; 2000US-243865P.
 XX
 XX (UTMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
 XX
 XX Baehrecke EH;
 XX
 XX WPI: 2002-479717/51.
 XX N-PSDB; AAD39238.
 XX
 XX Novel programmed cell death modulating proteins, useful for treating or
 XX preventing disorders associated with abnormal cell proliferation and
 XX apoptosis such as cancer, stroke, Parkinson's disease, myocardial
 XX infarction -
 XX

33XX
The

Disclosure; SEQ ID NO 3747; 21pp + Sequence Listing; English.
The invention relates to an isolated nucleic acid detection reagent

[illegible]

```

RESULT 9
ABB62231
ID ABB62231 standard; Protein; 2441 AA.
XX
AC ABB62231;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13485.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li FWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX N-PSDB; ABL06334.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 13485; 21pp + Sequence listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (AB557737-AB872072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 2441 AA;

Query Match 4.9%; Score 294; DB 22; Length 2441;
Best Local Similarity 19.8%; Pred. No. 8.8e-11;
Matches 246; Conservative 145; Mismatches 437; Indels 414; Gaps 48

7 EISLERVVEEGMRGROWKHVODKLTCSHLNIEOOPRIAGS----- 48
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1240 EAAANSVSTNSGCKKKKQKQAKNKVL---ISVEPPQPVITNSVLKGVCAKHPAVEVVKQ 1296
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49 -----EDEPSQYNHSSKEISQSNPNHCKTENHR-----LEQOHNGS 84
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1297 PPATQQAAPLRKQRLDVKKKEEPALKKKEKKNSSSSSSSKREKENLAPKEVALPAKQPASS 1356
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85 QLEEBEDSDENNQTHSDSRTPTPGATSPSPRRPRPIDMRPAKKNFCFN-GRLLTVNAG 143
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1357 SKLVSSSEASINIS--STAITSSANTTRKEVAPASQTASAT--TLNPAKRTVEDGWK 1411
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
144 KLVAESAA-----TATSSSTNS-NSHIHQH--DSDSNSASLPNHHISSSSS 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1412 EVAKKSSAQQTAAVAGSAGAPLPVATITATSIVQHNPNNHLLANSSNSSSLTTGTTTAA 1471
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
186 SNNNSGGR-----ARRIIAASARATPRAATPANSLELYKLLTORAPAKKTSMDSN-- 235

```

```

Db 1472 SVPEWTCKKVQVPAVNAISRVIGRGGSNINAIKATGAIHEVEKQKQNSERCITKGLTD 1531
Qy 236 AAQALQFLLA-----DENLINSLASQOOOQOOQOIAAVNP-----TTSVSA 281
Db 1532 ATKQHMILILKIDPDVILQMLPRINSSIKQASSGASITPWSVGTWDRFAAGVNAVY 1591
Qy 282 ISPALKDTPSPVDAPLDSKSPNSISGDPV-----SVRACATPTPSGRAYSEE 334
Db 1592 FSSAASIT---STSSSSASSTTPAGASYNAHKHQOQPOGVK-----PGRSTST-- 1641
Qy 335 DLSRALQDVANKLDARKSASQHEQRSILDNRLEFKMKHDOEQDHG-----DELED 387
Db 1642 -----VKNSGSGSTKVSASGSGSRGRAGSYLAQOQPGRSAGSGSSNGVIXKSES 1693
Qy 388 SDDAEAEVDNA-----STPVYPA--EFAPAQLRLKLSHSENG-----SD 427
Db 1694 SSKSLPAQKSTTLGKSTVSPGAQNPAAKAAIGQSSPKKAEAGATSAVVTSAAGRSSG 1753
Qy 428 LGEIDVGRSPKMGHPRACGNASANOAPASIPLDANVLTHTLMLAAGICAMPKLDQTV 487
Db 1754 VVAPRGKRPVAGCGPAAATASNVAQLGVSNGENI-----LAGPTG-----TFNV 1800
Qy 488 GDFIKGLLVANSGIMNEGILNLSAQSENSE-----GNASLLLQOQOHOQ 533
Db 1801 AD-VAANVAAAAAG-----AAATNSNVKPIAPLAPPSKRVGSPVQOQOHOQ 1849
Qy 534 HHQCH-----HQOQOQOQOHAAYHRRLPKSSTPTNSLDPNDASEPI 577
Db 1850 QOQOQOLPOPAPVPGPOQOOLQOQOQO-----APQOQOQOQOQOQOQOTSQOQ 1901
Qy 578 LKIPSFVKVSGPSSSSSLPGGLVGGHHHPLNNNSLSTSNNSHNSGHRNRSRPSHA 637
Db 1902 I-----NTNL-----1907
Qy 638 SPMILAAVAAQGYSAAGNSLLTSSSSSIQKMAASNIQROIENGESLRNGVSDCSSNN 697
Db 1908 NDLMASAA-----NTTSDSFSAQLAKLSAAYSLFSDVQOQOMKLG-D-PIG 1955
Qy 698 GSSSSLYKKKPSISVAKIIGCDTIRFGASPVLS-----QCHSHAHLT-----HQ 744
Db 1956 GGAGVVGGLPQADASKAP-----GYNRITLSSPVGSSKASSNHSSTSPVGNVIOQ 2007
Qy 745 QOQOQOLSAQEA-----GKGTPRKGRKYRNYRDSLVEAVKAVQR-----GEM 787
Db 2008 QOQOQOSSQALNITTSGBRGPATAPARS-----FMVAGNEBNPAVQGPBMNGTGLGET 2062
Qy 788 SVHRAGSYYGVPHSTLEYKVKERHLMRPKREPKQPDVLGLTGPANKLQDLKLAGPHG 847
Db 2063 APASHPGVIKPPTATVPI---QRHVPMPISAPBEAGAPPTFGAIG-----SNPAS 2108
Qy 848 GSKLSNALKNONNAAAAAATAATPAGLPLFEA-----GPO-- 891
Db 2109 G-----NNSHAAQAAAAAASAMIDROQONLQNLQTLQNLQNVGASQOQOPOQ 2158
Qy 892 -----ALSFQNMFWPTQNA-----TNAYGLDFNRITTEAMR 922
Db 2159 LNYPMWDTSSFIVDANNVLRNLRVIFPQGNTKPPQPPGGTQGVNG-----GNGR 2212
Qy 923 NPOASNHGKMSAODMVENYDGI-----RKTLOASBGNSSAAGNGSNGNGHG 975
Db 2213 QPPGT---GAROPGAAARWYGCTLEYPSYGRDMLHLENGAGMAGGSPSAMS PMHD 2269
Qy 976 HGHGHALLDGLLVKTPPLFTNHRNDYATCSSAGSEV 1017
Db 2270 D-----IRKMPRIPTERASWKNYNNFVGGPSL 2298

```

RESULT 10
 ABB61959 standard; Protein; 1436 AA.
 ID ABB61959
 AC ABB61959;

```

XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 12669.
XX DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX KW Drosophila melanogaster.
XX OS
XX WO200171042-A2.
XX PN
XX 27-SEP-2001.
XX PD
XX 23-MAR-2001; 2001WO-US09231.
XX PE
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA
XX (PEKE ) PE CORP NY.
XX PI
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL06062.
XX DX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS
XX Disclosure; SEQ ID NO 12669; 21pp + Sequence Listing; English.
XX CC
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AB57373-AB572072).
XX CC
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX Sequence 1436 AA;
XX SQ

```

Query Match 4.9%; Score 293; DB 22; Length 1436;
 Best Local Similarity 20.1%; Pred. No. 4.9e-11;
 Matches 232; Conservative 132; Mismatches 350; Indels 440; Gaps 48;

```

Qy 206 TPAATPANSLELYKLLTORAAK-----MTSMDMAAQLAQFSLADFNLSASQO 259
Db 415 KTKSGSSAGSLQROQALRGAVRGAMOCIFESHGHPAQLA-----NCFATPRELQ 466
Qy 260 QOQOQOIASAVTPTTSEVA-----AAISPALDTPS--PSVDAPLDLSKSPNS 308
Db 467 QLRQDMADNL-----ISASIFLRPLCPALISPLSLENTSELSARATRLTLY----- 515
Qy 309 SISGDVSVRACATPTPSGRAYSEEDLSRALQDVANK-----LDARKSASQHEQR 361
Db 516 -----AKTLQTLAFT---RFQGENFMELNLFLEQEARMOQFLEITSTREHPAPD 566
Qy 362 SIID-----NRLFKMKH-----DOEQDHGDELESDNDAEAVDSNASTPVY 405
Db 567 SILDMAGYIDGQKSLSHLSLSLAKLPPARQHELDPLQIHLDETSRAEHGKTL- 625
Qy 406 PAEFARQLRLKLSHSEH-----NGSDLGEDVDGRSPKMGHPRACGNASANOAPAS 457
Db 626 PGYLPATSSSTHSLASENGEARNPGSSGSHAGNSEQLLPQSO-----LAQOQHAIVS 679
Qy 458 IPLDANVLTHTLMLAAGICAMPKLDQTVGDFIKGLLVANSGIMNEGILNLSAQSEN 517
Db 680 KPLSAA-----RGI-----MRGVLTPNS---LEK---NIFRYNDPT 709

```



```

Db      930  AAGNSLNAPIVNOQLPEIPTNQOOSKISAPDLCA-ASLGGVYRSIQRP-----QTOKTL 981
QY      472  AAGIGAMEKLDIETQVGDPIKGLLVAANGGIMNEDLMLLSAQENSNGNASLLDQOOOH 531
Db      982  QPAIISSTQGPPTLPKPKRHKK---SNGGALDSSNAHMQTLPPKS---ASLRQOQOQO 1035
QY      532  QOOHHO-----OH-----HOOQO 543
Db      1036  QOOOERDRERERERPRPRAETLDNARSYIEQOQOQVPSHLITSSNMKKOHSDYSSYHQQO 1095
QY      544  QOOHAAARHRLPKSETPETNSLDLPNDASEPILKISFKYSGPASSSSLPGLVGGH 603
Db      1096  QOOAFYOROHNNFFSSKOOQOQOQOQOQALF-YRSLKGERVSTGS---GGVTGG- 1156
QY      604  HHPILNNNSLTSNNSSNHSN-----SHR-----NGSNRSPH----- 635
Db      1151  -----GAAISNOSDLXSNQAPVPLPPKKRORERENLREBOAPCNANLCEHEXD 1207
QY      636  ----SASPMLAAVAAGGVSAGNSLITSSSSI---OKMAS-----NIGROINE 678
Db      1203  EFYPAYEMHVAAGGTGPTGGAVATSSAGKGGQGGORATPFVADARVDELRLNLR 1266
QY      679  QSGQE---SLRNGV-----SD-----CSSNGGSSSLGYKPRS 709
Db      1263  RSQOQAAESKNGGAGAPIVSQDHHRSHDRERSRDHRIASVACEEVN--SSALCSAGSM 1322
QY      710  ISVAIIIGTDRSPRCASPNLLSQOHNHSAHNLTHQOQOQOQSLAQALGKTPRKPKX 766
Db      1321  SSMLOPANGTGHRS-----SHOHORDHNAARDQOQOQOQOQOQO-----RERREK 1365

```

	RESULT 12
ABB61328	
ID	ABB61328 standard; Protein; 1409 AA.
XX	
AC	ABB61328;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 10776.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
KW	
XX	Drosophila melanogaster.
OS	
EN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656660/75.
DR	N-PSDB; ABL05431.
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
XX	
PS	Disclosure; SEQ ID NO 10776; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of
CC	

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB120511), expressed DNA
CC sequences (AB101840-AB161175) and the encoded proteins
CC (AB577737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Query Match	4.7%;	Score 285.5;	DB 22;	Length 1409;
Best Local Similarity	19.9%;	Pred. No. 1.6e-10;		
Matches 200;	Conservative 144;	Mismatches 337;	Indels 323;	Gaps 43;

OY		131	CVNBRLLTVNAOGLVNESAAATSTSSGTSMSHIIHQHSDSDNSASALPHNHSISSSSSNNNS	190
Db		166	CLN-----VLVOHGVSVDYNKSSSGGRHKSSQOOQLHQHQOQQOQQOQ--QOHLSGSNGSNNS	220
OY		191	SGNBARHTAASAARATPAATAATPANSLELYLTLTORAKMTSMDSMAOLAQ----	FSLI 245
Db		221	SKOTSR-----SNTIKSKSSSTLSADVPFYLHPALFGSGSGSGMGJGMCMGMGMMSG	276
OY		246	ADEFULINS--LASGOOQOQOQOQOQIASANTPTTSESAAISPALKDTPSPSYDA-----	296
Db		277	LGISRKNSDALYSQOSKSSEKLNG-----SSLIAGNPGNSSGGCGVG	323
OY		297	-----PLD-----LSXKPSSNISGDVKSVRACTPPSGRAYSEDL	336
Db		324	GCGSGGALLLPNDGLVYPMPNMGMYNTPSPNGSIG-----ESFF	363
OY		337	SRLAQDVANKL-----DARKSAGSHHOKRSLDNRLFKMKHND----QEODHDEL	385
Db		364	LHPQDITIYNRVRDLFVDSGCCSSVKKGHGSHA---HOLLQSXAGNMVTIQADVHSSSG	420
OY		386	EDSNDDAEAVDSNASTPVYPAEFAQAOLKLS---HLSEHNGBDLDEVDRCGPKMGRH	442
Db		421	AGSGSDSVSISSSFNSPK-----QOOQLKSQSPNRRGSSSNISASNINK-----	466
OY		443	PACGNASANOCAPASIPLDANVLHTMLAIGAMPKLDETQTGVDFIGLVLVANGGI	502
Db		467	-----NNYKAHHKGKPSN-----QONCGD	484
OY		503	MN--EGLNLITLASQENSGNASILLQOQHQQH-N-QOHQOQOQOQHVAAVRYHRILPKSET	560
Db		485	HIDVEDITLVREBSKHQHQOQOQHOLQOQOOLQHOATQOQOQOQOQLOHKIVAGSRS	544
OY		561	PETNSSLPDPDASEDFLIKIPRFKVSQPASSSSLSPGCVL---GGHHPLNNNSLSISN	617
Db		545	RDSGSH-----SRASABSTRSTDIVLOSNNH--LNKNKNMNNS	583
OY		618	NSNHSS-----NSHRNGSNRS-----PHASPMILA--AVAQGYSAGNSLLTS---	659
Db		584	MNNSSSNIAOSSSSNNNNNSLNRNKSHBIIGLHSKEYESSLEDNYNSAKRVMLKQDLN	643
OY		660	-----SSSIQKMAMSNIOQINEOGSQBELNGVNUDCSSNNGSSSLGYKKPIS-VAKI	715
Db		644	GCISDPTDYEVSCPREDVAERTKOTHKMSMLRNULDASSNNNTSGSIN---NISNIGM	699
OY		716	IGTDTSRFGASPULISOQHSAHIILTHQOQOQOLSAOELGKTGRPKGRKIRNYDRDSL	775
Db		700	NGGQOSSR-----NLKRVSAPRMONTL--AVUNGPRP-----PPL	732
OY		776	VEAWKA---VORGEMSHRAGSYUGVPHSLFLKVKEKRHLMRPKRPKFORPLUGLTGP	832
Db		733	PPPLRAAVQOQRNNLUSDOPNSWTG-----NNGITYKANFGPNQ-----GP	773
OY		833	ANKLQDLKLKAGPHGSKLSNALKNONNOAAAAAAAAAAAAAATTNGLLPLFEAGPOA	892
Db		774	TN-----ANGTCGAA-----PPVPARPVPATEAVDSDEGLEV-----VEE	809
OY		893	LSFOPMFMFOQTATNAUYGLDENRITPEAMKNPOASHHGIMKSAQDMVENVYUOIIRKTL	952
Db		810	PSLRPSELVGNHRTMTSTISANKKALL--NASTINGSSIJAASDDS--QSRYGGSYA--	864

DR WPI: 2001-656660/75.
 DR N-PSDB; ABL09602.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 23289; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences and the encoded proteins
 CC (AB57737-AB572072).
 CC (AB57737-AB572072) and the encoded proteins
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 1416 AA;

Query Match 4.7%; Score 283; DB 22; Length 1416;
 Best Local Similarity 18.7%; Pred. No. 2.4e-10;
 Matches 258; Conservative 150; Mismatches 469; Indels 504; Gaps 53;

QY 20 RROMGHYDOKLTCSH-----LNIEEQPIAINGSEDEPQVYHSSKEISQSNPHNCKTEN 74
 Db 134 QQQQQHQQQRRATLQHPRLPNSNIKQASVKLIEGSDPKDTRNH--KKYVHKKAKKRSTTG 191
 QY 75 HRLQOHNQSQLEBEDSENNQTSHTDSRTPTGATSTPSPPEPIDMPPAKCNFCVNG 134
 Db 192 HHRHQOQQQQQQPHQQOQQQQ-----QQTLPTRSAQOQQOQQQLQQRHS----- 238
 QY 135 RLLTVNAGKLVASAAATRTSSSTGSHIHQSDSNSGASLPHHISSSSSNNNSGGR 194
 Db 239 ---TYN-----NNNGNSRKDKNSQYNSDSAP-----KSSDDE 269
 QY 195 ARHIAASARATPAATPANSLELYKLTQRAAKMTSMQMAQLAQ----- 241
 Db 270 ADIDEVEEQPLPVQAPVQS-----ASQPAAGLSPDSISDDVRLPOLQSPKRL 323
 QY 242 -----FSLP----- 249
 Db 324 TVAPAEHNKQPSPYYSDLTKSRDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 383
 QY 250 L--INSLASQOQQQQOQIASAIVPTTSEVSAAIAPALKOTPPSPVDAPLDLSKPSBN 307
 Db 384 LKCRGSTASEPQQTQTQTLGGYRKSSSLDVPPEEQPEAESHHEEDPTGDSQSTPKRY 443
 QY 308 SSISGDVSVRCATPTSGRRAYSEEDLSRALQDVVANKLDARKSASQHEQRIIDNR 367
 Db 444 SFTBEGVHIIR-CETPSTS--TSESDSCSECLK-----REWHARALALVKK 487
 QY 368 LFKMGHNOEQDHDDELE--DSNDDAEAVDSNASTFVVPAPAPARQLRKL----- 417
 Db 488 TCNVQ--FRNQNPTSEIQLQLOHCDAGEGELKKCCPPAPAE--NQSQVQMPRHLLG 544
 QY 418 -----SHLSEHNGSDIGEDVD--RGSPPKGRHP-----ACGNASANO 452
 Db 545 PAAVGCACITSAAPASLAGRSEIGDDEEYFPRRSIFVHPGVHECIEC--AQP 600
 QY 453 GAPASIPLDANVLTMLMAG-----IGAMK----- 480
 Db 601 AOPATVP--APTLSNPISLDIYDDEVDGDEDEDEDEDSRRRQIYETAFDCKI 658
 QY 481 -----LDETQVGFQFIKGLVANGGIMNEGLMLLSAQ-----ENSGNASIL 525
 Db 659 AKSDDDLDEVRITNCSVLQLSLNNG--GDIKIVTSSSTRAECKRAKSKQA--- 711
 QY 526 LQQQOQHQQHQQH-----QQQQQQH-----VAAVRRHLP 556

Db 712 LQETGEAIVOVHNEVELKQLQSQVDQONQONQSQOQHGVSATSGASASTOQLP 771
 QY 557 -KSERPEINSSLD-----PNDASEDPI-----LKIP-----SFK 584
 Db 772 VRGTPSPSTAPLPMKPFKGNDRYLMNSTRSAPNLPAPNAPRLRLDLPLOSMRHQ 831
 QY 585 VSGPSSSSLSPPGIVGHHHPHNNNSLSI----- 615
 Db 832 CGGSNDNSLTDNAV--NPTSTNSNMAASASASASVSATGSAAPRRPSFVLESGRV 888
 QY 616 -----SNNSHNSNGHRNCSNTPSPASPMLAAYAQGS--YSAGNSLITS 659
 Db 889 LELRSAGHOGQSTHHHHHHHHHAPGSRNRHGO--OGQTTGGHNYSTESIATS 945
 QY 660 SSSSIQKWMASNIQROIWESGQSLRNGVSDCSNNNGSSSIQYKKPSTIVAKITGT 719
 Db 946 SSGGMEBLRSTSEGNRTSSBSRSHSSLSHSSBSGSGSG-----SGGS 994
 QY 720 DTS-RFGASPMILSQOHSJAHILTHQOQQOQLSAQELGKGTBRPKGKYRNYDRDSLVEA 778
 Db 995 SVAYPLRAPLILSKHLILSPIDKSSQEPASAS----- 1031
 QY 779 VKAVORGEMSVHRAAGSYGVPHSTLEYKVKERHLMRPKREPKQPDVLGLTPANKQL 838
 Db 1032 -----EQSQPSQO-----LAPWASQED 1049
 QY 839 DKLKAGPHGSKLS-NLAKNQ--NNQAAAAAATAATNGLKLPFENG-PQAL 893
 Db 1050 DSTTSGADAGNOOVNASKPRSNRGAPNKAQLQLADELIGSDSGISLHREDEGKPOL 1109
 QY 894 SFQPMFMPQTNATNAYGLDFNRITTEAMRNPOASNHHGLMSAQDWMENVYDGIIRKTLQ 953
 Db 1110 ALQ-----RLT--LPKQLIGABETTTSA-----T 1132
 QY 954 ASEGNGSAGNGSNGNGHGHGHGHALLDQLVYKTPLPF-----TNRNNDVAT 1008
 Db 1133 ASGAGSSATGSSGAGSGSVVASCSTGIQDPL--RDLPPMPYLRRKRTLQCEAA 1189
 QY 1009 CSSAGSEVYKRSAGSPMGVADIK--RERLASDSSGSSDEHSASHHNNNSDLAHNKK 1066
 Db 1190 CTSGSATSVDIGELPF--DMPKLRRLRANQ-----BEINNL--LWHSTESS 1232
 QY 1067 GGGGGGGNGQTNNGRSGSRMTSR-----DSEETDASSFKSGENGQONHKMDL 1116
 Db 1233 GTSQASSHSNMRDQKWSKIDTALPQNLTNLNESHQATKQGSIDLRLNKKELNM 1292
 QY 1117 N 1117
 Db 1293 N 1293

RESULT 15
 ABB60091
 ID: ABB60091 standard; Protein, 1741 AA.
 XX
 AC ABB60091;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE *Drosophila melanogaster* polypeptide SEQ ID NO 7065.
 XX
 KW *Drosophila*; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS *Drosophila melanogaster*.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:26:02 : Search time 32 Seconds
(without alignments)
1540.380 Million cell updates/sec

Title: US-10-016-768a-10

Perfect score: 6030

Sequence: 1 MHISSEYISLSEVAECMGR.....ILHEKLAQIKAEVDQADQL 1165

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	252	4.2	2137	4 US-09-134-001C-4463	Sequence 4463, App
2	243	4.0	1507	4 US-09-914-259-37	Sequence 37, Appl
3	231	3.8	1093	3 US-08-545-860D-55	Sequence 55, Appl
4	231	3.8	1093	5 PCT-US94-04496-55	Sequence 55, Appl
5	229.5	3.8	2843	1 US-07-741-940-2	Sequence 2, Appl
6	229.5	3.8	2843	1 US-08-289-548A-2	Sequence 2, Appl
7	229.5	3.8	2843	1 US-08-452-654-2	Sequence 2, Appl
8	229.5	3.8	2843	1 US-08-370-235A-2	Sequence 2, Appl
9	229.5	3.8	2843	1 US-08-449-731-2	Sequence 2, Appl
10	223	3.7	1402	4 US-09-125-635-12	Sequence 12, Appl
11	221.5	3.7	2842	1 US-07-741-940-7	Sequence 7, Appl
12	221.5	3.7	2842	1 US-08-289-548A-7	Sequence 7, Appl
13	221.5	3.7	2842	1 US-08-452-654-7	Sequence 7, Appl
14	221.5	3.7	2842	1 US-08-449-731-7	Sequence 7, Appl
15	221.5	3.7	2843	1 US-08-452-655B-2	Sequence 2, Appl
16	221.5	3.7	2843	1 US-08-452-655B-7	Sequence 7, Appl
17	221.5	3.7	2843	1 US-08-450-582-2	Sequence 2, Appl
18	221.5	3.7	2843	3 US-08-450-582-7	Sequence 7, Appl
19	221.5	3.7	2973	2 US-08-821-355A-7	Sequence 7, Appl
20	221.5	3.7	2973	2 US-09-003-687A-7	Sequence 7, Appl
21	221.5	3.7	2973	2 US-09-136-605-7	Sequence 7, Appl
22	218	3.6	10182	4 US-09-134-001C-3159	Sequence 3159, App
23	214	3.5	1088	4 US-09-233-857-13	Sequence 13, Appl
24	213	3.5	1099	4 US-09-442-100-2	Sequence 2, Appl
25	213	3.5	1099	4 US-08-939-106-2	Sequence 2, Appl
26	211.5	3.5	1752	4 US-09-556-877-180	Sequence 180, App
27	211.5	3.5	1752	4 US-09-620-412C-180	Sequence 180, App

28	211.5	3.5	1752	4 US-09-598-419-180	Sequence 180, App
29	211	3.5	2703	1 US-08-185-432-19	Sequence 19, Appl
30	211	3.5	2703	4 US-08-899-232-4	Sequence 4, Appl
31	204.5	3.4	1581	3 US-09-110-517-2	Sequence 2, Appl
32	197	3.3	903	2 US-08-853-310-2	Sequence 2, Appl
33	193.5	3.2	1420	4 US-09-125-635-4	Sequence 4, Appl
34	193.5	3.2	1664	1 US-09-599-652-2	Sequence 2, Appl
35	193.5	3.2	1664	2 US-08-642-846-2	Sequence 2, Appl
36	193.5	3.2	1664	4 US-09-264-604-2	Sequence 2, Appl
37	186.5	3.1	951	4 US-09-125-635-8	Sequence 8, Appl
38	182.5	3.0	3969	4 US-08-061-376-5	Sequence 5, Appl
39	182	3.0	842	5 PCT-US96-02311-15	Sequence 15, Appl
40	182	3.0	1326	4 US-09-147-236-5	Sequence 5, Appl
41	182	3.0	1326	4 US-09-522-474-5	Sequence 5, Appl
42	180.5	3.0	1464	3 US-08-891-640-2	Sequence 2, Appl
43	179.5	3.0	1257	3 US-09-220-641-3	Sequence 3, Appl
44	178.5	3.0	760	1 US-08-195-152-2	Sequence 2, Appl
45	178.5	3.0	1776	4 US-09-556-877-179	Sequence 179, App

ALIGNMENTS

RESULT 1

US-09-134-001C-4463
Sequence 4463, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4463

LENGTH: 2137

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4463

Query Match

Best Local Similarity 17.1%; Pred. No. 2.5e-09;

Matches 201; Conservative 208; Mismatches 539; Indels 230; Gaps 33;

QY	51	EPSQVNHSSKESISQ---	NPNHCKTEHNRLEQOHNGSQLBE-----	EDSENNOTSHD- 100
DB	852	QASESASTSKSLESASTSDSAS	ESKSESTSKSLESASTSDSASVSTSESA 911	
QY	101	SSRTPTPATSPSPPEPPIDMR	PAKCNFCVNGRLVLVNAOGKLVASAAATATSSSTSN 160	
DB	912	STSTVSSGTSSTSTSTSTST	SDSASIKABESA-----	STSKLSASVSTSTSDAST 965
QY	161	SHIHQDSDNS-SASLPHNIS---	SSSSNNNSGNRAHIAASARATPAATPAN 214	
DB	966	S---TSVSDNSASTSKSST	SVSDSTSTSDSASTSESSSDASVSTSESA 1022	
QY	215	SLELYKLITQRAKKTSMDSMA	QAQPSLADFNVLNSLASQOQOQOQOQIAVPTPT 274	
DB	1023	VSDSTSTSTSDSASVSESES	-----	NSKSTSESTSTSGSTSAST 1067
QY	275	SEVSAATSPALUKPSPSPVAP	LDLSKPPNSISGDVSVACATPPSGRAYSEE 334	
DB	1068	SDSASTSESESDSTSL---	SESTSTSLSGST---	ASTSDASTSTSES 1114
QY	335	D-----LSBALDVPVANKL	DARKSASQHEQRSIIDLNRLLFKKKHHDOEDHGDDELED 387	
DB	1115	DSTSESTSLSTSLSTSVSD	STSASTSESA-----	STSTSESENSASTSLSG 1161

```

OY      368 S-----NDPAEAEVDSNMSTPVYPPEAPFAQJRKLSHSEHGSDLGEDVDGRGPKMRGH 442
Db      1162 SLTSTISDSTSTSTSDSASTSTSEBSDBTS---TSTSESTSTSLDSTSTSTSESAST 1217
OY      443 PACGNASANOAGPASIPLDANVLTLTMLAAGIGAMPKLDJETQYDFIKGLLVANSOGGI 502
Db      1218 STSESDSTSEESTSTSEESTSTSVSDST-----SASTSDASNTSTVSJD-----SSASTS 1266
OY      503 MNEGLINLISAOENSGNASILLLOOOQHQQHHQQHQQHQQHQQHQAAYRHRLPKSETPE 562
Db      1267 ISESLSTSVSDSTSTSTSDSA-----STSTSE 1293
OY      563 TNSSLDPMDASDPLLKIPSKFVGSPASSSTSPG-----GLVGGHHHPILNNNSL 613
Db      1294 SDSTSESTSLSESTSTSVSDSTSTASDTSDASTSTSEBSDSASTSLSGSTSTSLSTSTST 1353
OY      614 SISNNSMNS---SNRGVNGSRNRPSPASPMFLAAAVAGGYASGNSLLTSSSSSIGOMMAS 670
Db      1354 STSDASASTSTSBDEBRASSTSLSGSTSTSLDSTSTSTSDASTSTSVSDNSASTSLSG 1413
OY      671 NIQRINFQSGES--LRNGNVDCSSNNGGSSSLGYKKPSISVAKIICGTDTSPFGASP 728
Db      1414 SLSTSVSDSTSTSTSDSASASTSESDSERASTSLGSTSTSTSIDST---STSDSASTS 1470
OY      729 NLLSQOHHSAMHLTHQOOOQOLQSAEALCKGTRPRKXRYN-DEDL---VEVYXAVQ 783
Db      1471 TSVSSNSTS---TSTSESLSTSVSDSTSTSTSDASTSTSVSDSDASTSSSESVSTSD 1527
OY      784 RGENSVHAGSYGVPHSTLEKYVERHMLPRKREPKQPVLGLT-----830
Db      1528 SESTSTSTSDS-----ASTSTSVSESN-----STSTSLSGSTSTSVSDSTSTSTS 1572
OY      831 --GPANKQLDKLKACPHGGSKLNSALKNNQNNAAAAAAAAAAAAAATPNGKLPLFEA 888
Db      1573 DSASASTSESDSDASTSSSESVSTSV--SDSTASSTSEASTSTSVSDSNSASTSLSES 1630
OY      889 GPQALSFPNNFWPTTNATNAVGLDPNKITEAMRPQASNHGMLKMSAQDMVENYDDGIT 948
Db      1631 TSTSLDSTSMSTDSDASTSTSESDSDASTSLSDSTST-----1670
OY      949 KRTLOASGNGSAAAGNGSNGNGHGHGHGHALLD--QLLYVKTPLPFTNRHNDYA 1006
Db      1671 -----VSESTSTSTSTSVASANTSTSTSLSDRSKTSLSTSTSTSESGSTSESDSDSA 1723
OY      1007 AT-----CSSASGEYSVRKSGSPMGNYADIKERLS-----ADSGSSGDEBHS 1048
Db      1726 STSLSESTSTSLDSTSTSTSDASTSMVSVDNSNASTSLSDSTSTSVSDTSASTSSA 1765
OY      1049 ASHINNNSDLAHANNKKGSGGGCGGCGNQCTNGNGSGSMWTSRDSETDASSFKGECNGCQ 1108
Db      1786 STSTRESSTASTSLSESTSTSVSDSTSTSTSDASTSESDSNSSESTSLSESTSTSV 1845
OY      1109 QNHKKMDLNGSSSSSHIKCEBEATGHSFGCHHTSI 1146
Db      1846 SDSTASASTSASTSTSVSDSNSAST--LSGSTSTSV 1881

RESULT 2
US-09-914-259-37
; Sequence 37, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914.259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 1507

```

[illegible]


```

OY 634 PHSAPMLAAVAAGGYASAGNSLTLSSSSSIQKMMAS-----NIQPINQSGQESLR 686
Db 830 FHSTRPPL-----PLQQSPPTTLALPGAPAPLPQPPQNGLGRAPGAAGI- 875
OY 687 NGNVSDCSNNGSSSLGYKKPSISVAKIIG-----TDSRFGASPNL-----LS 732
Db 876 -GAMMAEGLLGLAGSG-----GLPLNGILGLINGMAAHPNPSLSQAGAPTLQLPGCLN 930
OY 733 QOHSAHLTHQOQOQOOLSADQALCKGTRPKRKYRNYDRDSLVAEYKAVQCGEMSVHRA 792
Db 931 SLTEGQRHLHQQEQQLQLQLQL-----A 955
OY 793 GSYVYVPHSTLEYKY-----KERHLMRPKREPKQPDVLGLTGPAKLIQLDKLKAGPHG 847
Db 956 SPQLTPEHQTVVYQMIQIQIKRELFQRIQMAQSGQLPMASLLAGSSTPL-----LSAGTPG 1011
OY 848 -----GSKLSNALKNQNNQAAAAAATAATTPNGKLPLFEAGPQALS 894
Db 1012 LLPTASAPLLPAGALVAPSLGNNTSLMAAAAAAAAVAA-----GGPPLYLT 1056
OY 895 FQPNMEWPQTNATNAYG 911
Db 1059 AQTNPFLSLSGAEGSGG 1075

```

RESULT 4

```

PCT-US94-04496-55
: Sequence 55, Application PC/TUS940496
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
: NUMBER OF SEQUENCES: 86
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
: ADDRESSEE: Norris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TUD-1242
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ. ID NO. 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1093 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US94-04496-55

```

Query Match	3.8%;	Score 231;	DB 5;	Length 1093;
Best Local Similarity	19.9%;	Pred. No. 3.3e-08;		
Matches 206;	Conservative 116;	Mismatches 373;	Indels 342;	Gaps 40

OY		19	GRQWKHYODKLTGSHNIEQGCI-----AIAEGEDBPQQYNHSSXKISQGN-	66
Dd		237	GQKSRDKDEKLKKQHKKRPBSPSILTPVVPADKYSSASSSHHEASTOTTESSR	296
OY		67	-PNHKCTENHLLEOHNQSOLLEED-----SENNQTSHDSRTPTPGATST----	112
Dd		297	ESKGKSSHSBL--SHKQKLLSGKVYSFTASSSSSSSSSSGCFOPAVSLQSSPD	354
OY		113	-----PEPPPEPIDMRSAKCNFCVNRLTLTVNAQGLVAESA	150
Dd		355	FSAPFKLEQPEEDKYSKTAPAPAPAPSAPAEPPKADF-----EQVVVPSGF	403
OY		151	ATATSSSTSNHIIHQHDSDNSASALPHHHISSSSSSNNNSGNGRRHIAAASARATPAA	210
Dd		404	GPIMRFSTTS-----SSGARARPSPGDYKSPIVTGSAGAGNHKMPALSTATPVPADE	457
OY		211	TPANSLELYKLTGPAAKT-----SMDMAAQLOAFSLIADF	248
Dd		458	TPEGLKEKK--HKASKRSRHGPRKPSHKEGTGGAPAPLPS--AQLAGFTATA--	510
OY		249	NLIINSIASQQQQQQQQLASAIVTTTSEVSAALISPALKDTPSSVAPL-----DLSSKP	304
Dd		511	-----ASPFG-----GSLVSSGLGGLSRTTGPS-GSLIPSLSLSPPLGAGITYSNK	557
OY		305	SPNBSIGDVSVACACTPT-----PSGRAVEEDLSRALQDVANKLDARKASQNH	358
Dd		558	DPISHSGMLNAV--CSTPLSSSLGPPGTA-----LPRIISRPFSTLPPSSA----	605
OY		359	EQRSLIDNRLFPMKHNDQEDHDDDELBDSDNDAEAYDSNASITPVYPAEFARAQLKLS	418
Dd		606	--SISTTYVSTL-----AGSTFLSPSHIHFTGPMGVANPLLISQ	641
OY		419	HLSHNGSDJGEVDVDRGPKMGRPACNASANOCAPASIPLDANVLHTLMILAIGAM	478
Dd		642	AESHTEBEDLEDGFR-----CRGHSPOGSLSMSP-----ISL	676
OY		479	PXL-DETQTVGDFIKGLLVANGGIN--EGLNLISAOENSNGNALLLQQQHOQH	534
Dd		677	PALPDOTAS-----APCGGOQDPAPAGTTNMEOULEKQDGEAGNIVEMLKALH	727
OY		535	HQOHOOOQOQO--HVAAVRHRLP-----KSETPETNS-----SLDPNDAS	573
Dd		728	ALQXENORLEQILLTKKERLOILNVOLSVPFPALPALPANGVPCGYGLPROAGS	787
OY		574	EDPILKIPIKFVSGPASSSSLSPGALVGHHHPLNNNNSLSINNSHSSSHRNGBNS	633
Dd		788	SDSL-----STKSPFG-----KSLGLDNLSISTSEDDPHSGCPRSSSLS	829
OY		634	PHSASPMLAAVAOGGISAGNSLLTSSSSSIQKMMS-----NIQOIINEGQEBELR	686
Dd		830	FHSITPPPLP-----PLLQSPATLPLPALPAPADLPPOPONGLARAGAAGL-	875
OY		687	NGANTSDCSNNGSSSLGYKKPISVAKTIIG-----TDSRFCASPUL-----LS	732
Dd		876	-GAMPMAEGLLGLAGSG---GIPLINGLOGLANGAAPNAPASISOAGATTLOPGLTN	930
OY		733	QOHSASHHLTHOQOQOQOLISAQELGCKTRPKRGKYRYNDRDSLVEAVKANORGEMSVHRA	792
Dd		931	SLTEQORHLLODOEQOLOLOQLL-----A	955
OY		793	GSYYGVBPSTLEYVU-----KERILMAPRKREKROPDIVLGTGPANKLJDOKKAGPHG	847
Dd		956	SPQLTPEHQIVVMIOIQIQKRELORLOMANAGSQLPWASILLAGSSTPL-----LSAGTPG	1011
OY		848	-----GSKLSNALKNONNOAAAAAAAAAAAAATPENGKLPLEFAGPOLAS	894
Dd		1012	LLPTRASAPPELLPAGALVAPSLGNNTSLMAAAAAAAAAAAAAA-----GCPPLYLT	1058
OY		895	FQPMFMFOYTATNYAG	911
Dd		1059	AQTNPFSLSGAEGSG	1075

RESULT 5
US-07-741-940-2
Sequence 2, Application US/0741940
Patent No. 5352775
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-741-940-2

Query Match 3.8%; Score 229.5; DB 1; Length 2843;
Best Local Similarity 19.3%; Pred. No. 1.6e-07;
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY 2 HISSEISLIERVAEECMGRRO---MKGYODKLTCSHLNIE---EQQPIA---IAGS 48
DB 1054 HIIIEBI---KQSEKQOSNRQSTTYVYTESTDDKHLKQPHFGQECSPFRSGANGS 1110
QY 49 EDEPQOYHS-SKEISQSNPNHCKTENHRLLEQOHNGSQLLLEEDSENNQTSRTPPTP 107
DB 1111 ETRVQSNNGINQNSQS---LCQEDDYEDDKPTVYSERYSEEE---QHEEBERPTN 1161
QY 108 GA---TSTPSPPEPIDMRPSAKCNVCVNGRLITVNAQKLVASATATSSSTNSHH 164
DB 1162 YSIKNEERKRVHDQPIDY-----SLKVTDPISQKOSF 1195
QY 165 QHDSNSGASLPHHISSSSSSSNNNSGNRAR---HIAASARA---TPAATPANSLE 217
DB 1196 SFSKSSSGSSSTEHSSSSSESTSTSSNAKQONQHPSSASRSQPOKAAITCKVSSIN 1255
QY 218 LYKLLTORAAKMTSMDMAQLAQSFLADFNILNSLASQOQOQOQOQOIA-----S 268
DB 1256 QETIQYCVEDTPIPCFSCSSSLSSLSAEDICGNQTTQEADSANTLQIAEIKGKIGTRS 1315

QY 269 AVPTTSEVSAAPALKDTPPSPVDAPLDSLSPSSNSISGDKVSVACATPTPSGR 328
DB 1316 AEDP-VSEVPVAVSQHPRTK-----SRLOGSSLSSE-SARHVAVFPPGCA 1358
QY 329 RAYSEEDLSRALQDVANKLDARKSASQHEQRS-ILDNRLEFKMGHD-----Q 376
DB 1359 KSPSK-----SGAQPKSPPEHYQETPLMRSCSVSSLSLSPFSRSIASSVQ 1406
QY 377 EODHDG-----DELEDSDDAEVDSNASTVYVAEFARQALRLSHLSHNSDL 428
DB 1407 SEPCSGVSGIISPDLPS--PGQTMPPSSSKTPPPPTQATK-REV---PKMKAPT 1459
QY 429 GEDVDRGSPKGRHACGNASANOGAPASIPIDAVLH--TLMAAGGAMPKLDERTQ 486
DB 1460 AEKRESG-PKQ---AAVNAVQR---VQLPDAUTLHPATESTPDPGSCSSLSALS 1511
QY 487 VGDPIKGLLVANSQGINNEGILNLSASQENSNGASLLLOQOQHQQHHQOHHQOQOQO 546
DB 1512 DEPIQ-----KDVELRIMPVQENDNGN-----ETESQPKESNENQEKAEK 1555
QY 547 HVAAYRHRLPKSE-----TPETNSLDPNDA---SEDPIL 578
DB 1556 TIDSEKDLDDSDDDIELECIISAMPYKSRKKKPAQTASKLPPVARKPSQLPVY 1615
QY 579 K-IPSFKVSGPASSSLSPGC-----LVGGHHPLNNSLSISN-----NSNHSN 624
DB 1616 KILPSQNRLOPQKAVSFTPGDDMPRVYVEG--TPINSTATSLSDLTITESPDLAIGE 1673
QY 625 SHRNGSNRSPHASFPL-----AAVAQGYSA-----GNSLTTSSSSSIQ 665
DB 1674 GVRGAQSGEPEFKRPTITEGRSTDEAQGKTSSTVTELDONKAEEDDILAEICNSAMP 1733
QY 666 K-----MNASNIQIINEGSGQESLRNGVNSDCSNNNGSSSLCYKKPISIAKIGCT 719
DB 1734 KGSKHPRPVKKIMQVOOASASSAPKNQDL-----GKKKKPTSPVPIPON 1782
QY 720 DTSRGASPNLLSQOHHSAHLHTHQOQOQOQSLAQALCKGRPKRGKY-RVYDRSLVEA 778
DB 1783 TEYRTVRKADSNNLAEERFSDNKD--SKKNLKNNSKDFNDKLPNNEDR----- 1833
QY 779 VKAVORGEM---SVRAGSYGVPH-----STLEYKVERHLMRRKREPKPQPLV 827
DB 1834 ---VRGFAFDPSPHHYPIEGTFCFSRNSLSLDDDDDDVLSR----- 1876
QY 828 GLTGPAKQLQDKLKAGPHGSKLSNALKNQNNQAAAAAATAATPGLPLFE 887
DB 1877 ---EKAELRKAKENKESAKVTSHTELTSNQOSAKTKQAI-----KQPINR 1920
QY 888 AGPQALSFQPMNFWQTN-----ATNAVGLDPRNRTTEAMRN-POASNHGMLKMSAD 938
DB 1921 GQPKPI-LQKSTFPQSSKDIIDRGAAATDEKQNR---AIENTPVCFSHNSLSLSLD 1974
QY 939 MVE---NVYDGIIRRTLOASEGNGSAAGNGSNGNGHGHGHALLDOLLVKKPTL 995
DB 1975 IDQENNNENEPKKT-EPPOSQEP-----SKPQASGVAPKSFH-----VEDPIV 2019
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVKSGSPMG-----N 1026
DB 2020 CFSRNSLSLSLSIEDDLQECISSAMPKXKXKPSRLKGNDEKHSPRNNGILGEDTLTD 2079
QY 1027 YADIKR---ERLSDS-----GSSSDEH----- 1047
DB 2080 LKDIOFDPSEHGLSPDSENFPMKALQEGANSIVSSLHQAAAAACLSROASDSDSLTLK 2139
QY 1048 ---SASHINNNNSDLAHHNKNKSGGGGCGNGQOTGNGRSSRMTSRDSETPASSFK 1101
DB 2140 SGISLSPFHLTPDEEKPFTSNK-----GRILKPKGKSTLETIKSSES----- 2185
QY 1102 SGENGQONHKKMDLNGSSSSSHIKCE-----SEATGHHSPGHHTTS 1145
DB 2186 KGIGKGVKYSKL-ITGKVRNSSEISQMKOPLQANMPSISRGRTMIHPVRNNS 2240

RESULT 6
US-08-289-548A-2
Sequence 2, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-289-548A-2

Query Match 3.8%; Score 229.5; DB 1; Length 2843;
Best Local Similarity 19.3%; Pred. No. 1.6e-07;
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY 2 HISSVEISLIERVAECMGRQ--WKHYQDKLTCSHLNIE---EQQPIA-----IAGS 48
DB 1054 HIIIEDEI--KQSEOROSNQSTTYPVYTESFDKHLKQPHFGQOECVSPYRSGANGS 1110
QY 49 EDEPQOVNHS-SKETSSQSNPHNCKTENHRLBEOHNGSOLLEBEDENNOTSISDRTPLP 107
DB 1111 ETNRVGSNHNQINQVNSQ--LCQEDDYEDDKPTVYSERYSEEE-----QHEBERPTN 1161
QY 108 GA---TSTPSPPEPIDMPSPAKCNFCVNGRLITVNAQKLVAAESAATATSSSTNSHIH 164
DB 1162 YSIKNEERKHYDQIDY-----SLKTAITDIPSSQKQSF 1195
QY 165 QHDSQNSASLPHHIISSSSSSNNNSGGRAR---HIAASARA---TPAATPANSLE 217
DB 1196 SFSKSSSGGSKSTEHMSSSESTSTPSSVAKRQNLHPSSAQSRGQPGKATCVSSIN 1255
QY 218 LYKLTITRAAKTSHDSMAAQLAQFSLADENLINSLSAQOQOQOQOQOIA-----S 268
DB 1256 QETIQTVCVEDTPICFSSRCSSLSLSASADEIGCQNTTOEADSNANTLQIAETIKGKIGTRS 1315

QY 269 AVPTTSEVSAASIPALKDTPSPVDAPLDLSSKSPSSISGQDVKSVRACATPTPSGR 328
DB 1316 AEDP-VSEVPVAVSQHPRK-----SRLQSSLSSE--SARIKVAEPFPGA 1358
QY 329 RAYSEEDLSRALQDVANKLDARKASQHEORS-IIDNRLFKNKHD-----Q 376
DB 1359 KSPSK-----SGAQTPKSPPEHYQETPLMFSRCTSVSLDSFESRSIASSVQ 1406
QY 377 EGDHGC-----DELEDSNDDAEAVDSNASTPVVPAEFARQRLSLSHSNGSDL 428
DB 1407 SEPSCGMVSGIISDLPDS--PGQTMPFSKSTPPPPQTAQTK-REV---PRNKAFT 1459
QY 429 GEDVDRGSPKMGHPACCNASAGAPASIPLDANVLH--TLMAAGIGAMPKIDETOT 486
DB 1460 AEKRESG-PKQ---AAVNAAVQR---VQVLPAADTLHFATSTPDCGSCSSLSALSL 1511
QY 487 VGDFTKGLLVANSGCINNEGILNLISASQENSNGNASILLQOQHQQHQQHQQOQOQ 546
DB 1512 DEPFIO-----KVELIRIMPVOENDNGN---ETSEOPKESNENOKEAEK 1555
QY 547 HVAAYRHRLPKSE-----TPTNSLDPNDA---SEDPIL 578
DB 1556 TIDSEKLLDSDDDDEILEECTISAMPTGSKKKKPAQTAKLPPVARKPSQLPVY 1615
QY 579 K-IPSFKVCSPASSSSLSPGG-----LVGHHHPLNNNSLSISN-----NSNHSN 624
DB 1616 KLTPSQNRLOPQKIVSFTPGDDMPRYCVBE--TPINFTATSLSDLTIESPRELAAGE 1673
QY 625 SHRNGSNPSPHASPMI-----AAVAQGYSA-----GNSLITSSSSSIQ 665
DB 1674 GVRGAQSGEPEKEKDTIPTEGRSTDEAQGKTSVITPELDNKAEBEDILAEICINSMP 1733
QY 666 K-----MMAINIQIINEGCGESLNGNVSDDSSNNGSSSGYKKPSISVAKIIGCT 719
DB 1734 KKKHKKPRVKKINDQVQOASASSAPVKQLD-----GKKKKPTSPVKPIPON 1782
QY 720 DTSRFQSPNLISQOHSANHLTHQOQOQOQLSAEOALCKGTRPRKGY-RNYDRDSLVEA 778
DB 1783 TEYTRVAKNADSKNNLNAERVFSDNND--SKQNLKNNSKDNDKLPNNEDR----- 1833
QY 779 VKAVORGE--SVHRAGSYGVPH-----STLEYKVERHLMPRKRPEKPODLY 827
DB 1834 ---VRGSFAPDSPHHYPIEGTPYCFSRNDSLSTLFDDDDDVDLSR----- 1876
QY 828 GLTGPAKLOLDKIKAGHGGSKLSNALKNQNNQAAAAAATAATPNGLPLFE 887
DB 1877 -----EKAELRKAKENKESAKVTSHTELTSNQASANKTOAIA-----KQPINR 1920
QY 888 AGPALSFPQNMFWPQTV-----ATNAYGLDFNRITTEAMRN-POASNHHGLMKSAD 938
DB 1921 GQPKPI-LQKXSTPQSSKQIDPRGATDEKQNF-----AIENTPVCFHSNSSLSSLD 1974
QY 939 WVE---NVYDGIIRKTLQASEGNSAAGNSGNSNGHGHGHGHALDQLLYKKTPL 995
DB 1975 IDQENNNKENPIKET-EPPDSQGEPP-----SKPOASGYAPKPSFH-----VEDTPV 2019
QY 996 PETNHRN-----NDYATCSSAS-----GESVKSQSGSPMG-----N 1026
DB 2020 CFSRNSLSLSISIDSEDDLOECISAMPKKKPSRLKGDNKSPRNMGGITLGEDULTLD 2079
QY 1027 YADIKR-----ERLSADS-----GGSDEEH----- 1047
DB 2080 LKDIQRPSEHGLSPDSENFPMKALIQEGANSIVSLHQAALAAALSRQASDSISLSLK 2139
QY 1048 -----SASHINNNNSDLAHNKNKSGGGGGGNGQTNNGRSGSRMTSRDSETDASSFK 1101
DB 2140 SGISLSPFHLTPQEEKPFTSNK-----GRIILKPGKSTLETKLIESE----- 2185
QY 1102 SGENGQGNHGMMLNLNGSSSSSHIKCE-----SEATGHHSPEGHHTTS 1145
DB 2186 KGIGGKKVYKSL-ITGKVRNSSEISGOMKQPIQANMPISISRGRTMHIPIVRNSS 2240

RESULT 7
US-08-452-654-2
Sequence 2, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODNE, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-654-2

Query Match 3.8%; Score 229.5; DB 1; Length 2843;
Best Local Similarity 19.3%; Pred. No. 1.6e-07;
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY 2 HISSEISLAEVAEECMGRQ--WKHYDKLTCTSHLNI-----EQQPIA-----IAGS 48
DB 1054 HIIIEBEI--KQSEOROSHNOGTTPVYESTDDKHLKQPHFGQOEVSFYRSRGANGS 1110
QY 49 EEPSEQVYNS-SKETISQSNPHNCKTENHLEQOHNGSOLLEEDSENNOTSHDSRTPTP 107
DB 1111 ETNRVGSNNGINQNSQS--LCQEDDYDDKPTVNSERYSEE-----QHEEERPTN 1161
QY 108 GA---TSTPSPPEPIDMRPSAKCNFCVNGRLITVAQGLVAESNATATSSSTNSHIH 164
DB 1162 YSIKYNEXRHYVDQPIDY-----SLKYATIDIPSSQKQSF 1195
QY 165 QHDSQNSASLPHIHISSSSSNNNSGRRAR---HIAASARA---TPAATPANSIE 217
DB 1196 SFSKSSSGSSKTEHSSSSSENTSTPSSNAKRONOLHPSSAQRSGQOPKAAATCKVSSIN 1255

QY 218 LKLTLPQAAKWTSMDSMAAQLAQLFSLADPFLNLINSLASQOQOQOQOQOIA-----S 268
DB 1256 QETITVCVEDTPICFSSCSLSLSAEDELICGQTTQDEADSNANTLOIAELKIGITRS 1315
QY 269 AVTPTTSEVSAALISPAKDTSPSPVDAPLDSLSPSPNSSISGDKVSKVACATPTPSGR 328
DB 1316 AEDP-VSEVPVAVSQPRTK-----SSRLQSSLSSE--SARHKAVEFPSCA 1358
QY 329 RAYSEEDLSRALQDVNAKLARKASQHHQRS-ILDNRLFKMKHH-----Q 376
DB 1359 KSPSK-----SGAQTPKSPPEHYVQETPMLMSRCTSVSLDSFESRSIASSVQ 1406
QY 377 EODHDG-----DELESDNDAAEVDVSAFTVPYPAEPARAOLRLKLSLSENGSDL 428
DB 1407 SEPCGWSGIIISPDLDLS--PGOTMPPSSKTPPPPPQTAQTK-REV---PKNAPT 1459
QY 429 GEDVDKSPKXGRHPACGNASANGAPASIPLDANVLLH-TLMLAIGAMPKLDQTQ 486
DB 1460 AEKRESG-PKQ---AAVNAAVQR--VQVLPDADTLHFATSTPPDGFSCSSLSALSL 1511
QY 487 VGDFIKGLLVANSGGIMNEGILNLISASQENSNGASLLQOQOQHQQHQQOQOQO 546
DB 1512 DEPTIQ-----XVELRLIMPVQENDNGN-----ETESQPKESNENDEKAEK 1555
QY 547 HVAAYRRLPKSE-----TPETNSLDPNDA---SEDPIL 578
DB 1556 TIDSEKDLIDSDDDDIIEECIISAMPTYSRKGKPAOTASLPLPPVARKPSQLPVY 1615
QY 579 K-IPSFKVSQPASSSSLSPGG-----LVGHHPLNNNSLSISN-----NSNHSN 624
DB 1616 KLPFQNRLOQKQKVSFTPGDMPRVYVEG--PTINSTATSLDILTSPPNELAGE 1673
QY 625 SHRQNSNSPHSASAML-----AAVAQGYSA-----GNSLITSSSSSIQ 665
DB 1674 GVRGAQSGEPEKRDITPTBGRSTDEAQGTSVTIPELDNKAEBEIDIIECINSAMP 1733
QY 666 K-----WMAINIOQINEQSGQESLRNGVSDCSNNNGSSSGYKXKPSISVAKITIGT 719
DB 1734 KQKSHKPRVKKIMDOVQASASSAPKKNQD-----GKKKKPSPVKPIQON 1782
QY 720 DTSRPGASPNLISQOHSAMHLTHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 778
DB 1783 TEYRTRVAKNADSKNNLAERAVFSDNKD--SKQNLKNSKDFNDKLPNNEDR----- 1833
QY 779 VKAVDQGM---SVHRAGSYGVPR-----STLEKVKERHLMRKRKEPKQPDV 827
DB 1834 ---VRGSPAPDPSPHHTPIEGTPTCYFSRNDLSLDFDDDDVDLSR----- 1876
QY 828 GLTGPNKLOLDKLCAGPHGSKLSNALKNNQNNQAAAAAATAATPNGKLPUE 887
DB 1877 -----EKAELRKAKENKESAKVTSHTELTSNOQSAKKTQAI-----KQPINR 1920
QY 888 AGPQALSFQPMFMFQTN-----ATNAYGLDFNRIITEMARN-POASNHGMLKMSAD 938
DB 1921 GQPKPI-LQKSTFPQSSKDIIDRGAATDEKLQNR-----AIENTPVCFSHNSLSLSLD 1974
QY 939 MVE---NVYDGIIRKTLQASEGNGSAAAGNSGNGNGHGHGHGHALLDQLLYKKTPL 995
DB 1975 IDQENNNKNEPIKET-EPPDSQGEF-----SKPQASGYAPKSFH-----VEDTPV 2019
QY 996 PFTNHRN-----NDYATGSSAS-----GESVKSGSPMG-----N 1026
DB 2020 CFSRNSLSLSISDEDDLQECTISSAMPKPKKPSRLKGDNEKSPRMMGGILGEDTLTD 2079
QY 1027 YADIKR---ERLSADS-----GGSSDEEH----- 1047
DB 2080 LKDIQRPSSEHGLSPDSENFDMKALIQEGANSIVSLHQAAAAACLSRQASDDSDSILSLK 2139
QY 1048 -----SASHINNNSDLAHNKVSGGGGGGNGQNTGNKRSSAMTSRDSSEFDASSFK 1101
DB 2140 SGISLSPFHLTPQDEEPRFTSNK-----GPRILKPGKSTLETUKIESSE----- 2185
QY 1102 SGENGQOQNNHMMQDINGSSSSSHIKCE-----SEAATGHHSPGHHTTS 1145

Db 2186 KGKGGKVVYKSL-ITGKVRNSNSEISGOMKQPLQANMPSISRGRTMIHPGVRRSS 2240

RESULT 8

US-08-370-235A-2

; Sequence 2, Application US/08370235A

; Patent No. 5910418

; GENERAL INFORMATION:

; APPLICANT: VOGELSTEIN, BERT

; APPLICANT: KINZLER, KENNETH W.

; APPLICANT: HILL, DAVID E.

; APPLICANT: JOHNSON, KAREN A.

; TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING

; TITLE OF INVENTION: MUTATIONS IN THE APC GENE

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BANNER & WITCOFF, LTD.

; STREET: 1001 G STREET, N.W.

; CITY: WASHINGTON

; STATE: DC

; COUNTRY: US

; ZIP: 20001

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/370,235A

; FILING DATE: 01-JAN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: KAGAN, SARAH A.

; REGISTRATION NUMBER: 32,141

; REFERENCE/DOCKET NUMBER: 01107, 48688

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202 508 9100

; TELEFAX: 202 508 9299

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2843 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-370-235A-2

Query Match 3.8%; Score 229.5; DB 2; Length 2843;

Best Local Similarity 19.3%; Pred. No. 1.6e-07;

Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;

QY 2 HISSVEISLERVAECMGRQ---WKHYODKLTGSHLNE-----EQQPIA-----IAGS 48
Db 1054 HIIIEI---KQSEQRQSHNQSTTYPVYTESTDDHKLKQHPHGOECVSPYRSGANGS 1110
QY 49 EDEPSQVYNS-SKEISQSNPNHCKTENHRLQEOHNGSOLLEBEDENNQTSHTSSRTPTP 107
Db 1111 ETRVGSNHNGINQNSQS---LCQEDDYEDDKPTVYSEYSEB-----QHEDBERPTN 1161
QY 108 GA---TSTPSPPEPIIDMRPSAKCNFCVNGRLITVNAQGLVAESAAITTSSTNSHIH 164
Db 1162 YSIKYNKERRHVDQIDY-----SLKXATDIPSSQKQSF 1195
QY 165 QHDSQNSGASLPHHIISSSSSSNNNSNGNRAR---HTAAGARA---TPAATPANSLE 217
Db 1196 SFSKSSSGQSSKTEHMSSESTSTPSSNAKQONLHPSSAOSKSGQPKAATCVYSSTIN 1255
QY 218 LYKLTORAAKMTSDMSMAQLAFSLADFNLIINSLSAQOQOQOQOQOIA-----S 268
Db 1256 QETICTYCEDTPIPCFSSRSSLSLSADELICGNOTTOEASANTLOJAEIKKIGTRS 1315
QY 269 AVPTTSEVSAAIIPALKDTPSPVADADLISKSPSPNSISGCVKSVACATPTPSGR 328

Db 1316 AEDP-VSEVPVAVSQHPRTK-----SSRLQSSLSSE--SARRKAVEPBGCA 1358
QY 329 RAYSEEDLRALQDVANKLDARKSASQHEQRS-ILDNRLFPKKHND-----Q 376
Db 1359 KSPSK-----SGAQTPKSPPEHYQETPLMFRCRTSVSSLSDFSRSIASSVQ 1406
QY 377 EQDHDG-----DELEDSDDAEAEVDSNASTPVVYPAEFARAQLRKLSHLSEHNSDL 428
Db 1407 SEPCSGWVSGIISPSDLPDS--PGQWMPSSKTPPPPPQQAQTK-REV-----PKNKAPT 1459
QY 429 GEDVDKSGPKMGRBPACGNASANOAPASITLDANVLH--TMLAAGIGAMPULDEYOT 486
Db 1460 AEKESG-PKQ---AAVNAVQR---VQVLPDADTLHFATYSTPDGFCSSSLALS 1511
QY 487 VGDFIKGLLVANSGIMNEGILNLSASQESNGNASLILQOQOQHQQHQQHQQOQOQ 546
Db 1512 DEPTIQ-----KQVELRIMPVQENDNGN-----ETESQPKESNENQEKAK 1555
QY 547 HVAAYRRRLPKSE-----TPETNSSLDPNDA--SEDPIL 578
Db 1556 TIDSEKDLDDSDDDIEILEECIISAMPTKSRKKKPAQTASGLPPVARKPSQLPVY 1615
QY 579 K-IPSPKYSGPASSSSLSPGS-----LYGHHHPPLNNNSLSISN-----NSNNSN 624
Db 1616 KILPSQNLQOPQKVSFTPGDMPRVYCVES--TPINFSTATSLDLTTEPPELAGE 1673
QY 625 SHRNGSNRSPSASPML-----AAVAQGYSA-----GNSILTSSSSSIQ 665
Db 1674 GVRGAQSGEFKQDITPTEBRSTDEAOGKTSVTIPELDNNKAEEDILAEQINSAMP 1733
QY 666 K-----MMASNIQRIQNEQSGESLRNGVSDCSSNNGSSSLGYKKPSISVAKIIGCT 719
Db 1734 KGKSHKPPRVKIMDQVOAASASSAPNQLD-----GKKKKPTSPVKPIPON 1782
QY 720 DTSRFGASPNLISQOHSANHLTHQOQOQOQSLAEBALCKGRPRKGY-RVYDSDSLVEA 778
Db 1783 TEYTRVRKKNDSKNNLAERVFSDNKO--SKQONLKNNSKFNDKLPNNEDR----- 1833
QY 779 VKAVORCEM---SVFRAGSYGVPH-----STLEKYVERHLMRPRKREPPQPDV 827
Db 1834 ---VRGSFADSPHHHTPIGTPTCYCFSRNDSLSLDPDDDDVLSR----- 1876
QY 828 GLTGPANKLQDKLKAGPHGSKLSNALKNQNNQAAAAAATAATPNGLKLPLE 887
Db 1877 ---EXAEIRKAKENKESPAKVTSHTELTSNQSANKTOAIA-----KQPINR 1920
QY 888 AGPQALSTQPMFPMQTN-----ATNAYGLDPNRTITEMRN-POASNHGMLKSAQD 938
Db 1921 GQPKRI-LQXOSTPQSSKDIIDRGAATDEKLQNF---AIENTPVCFSHNSLSLSLD 1974
QY 939 MVE---NYVDGIIIRKTLQASEGNGSAAAGNSGNGHGHGHGHALLDQLVKTPL 995
Db 1975 IDQENNNKENPIKET--EPPOSQEP-----SKPAQSGVAPRSFH-----VEDTPV 2019
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVKSQSPMG-----N 1026
Db 2020 CFSRNSLSLSISIDSEDLIOECISSAMPKKKPSRLKGDNEKSPRMNGCILGEDTLTD 2079
QY 1027 YADIKR---ERLADS-----CGSSDEH----- 1047
Db 2080 LKDIORPDSEHGLPDSENFDPWKAIOEGANSIVSLHAAAAAACLSSQASDSISLSLK 2139
QY 1048 ---SASHINNNNSDLAHNKNKSGGGGGGNGQNTGNRSGSRMSTRDSEYDASFK 1101
Db 2140 SGISLGSFHLTPQOEKPTSNK-----GRILKPGKSLTFKXIESE----- 2185
QY 1102 SGENGGQGNKXMDLNGSSSSSHIKCE-----SEAAIGHSPGHHTTS 1145
Db 2186 KGKGGKVVYKSL-ITGKVRNSNSEISGOMKQPLQANMPSISRGRTMIHPGVRRSS 2240

RESULT 9

US-08-449-731-2

Sequence 2, Application US/08449731
Patent No. 6413727
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
ANAND, RAKESH
CARLSON, MARY
GRODEN, JOANNA
HEDGE, PHILIP J.
JOSELYN, GEOFF
KINZLER, KENNETH
MARKHAM, ALEXANDER F.
NAKAMURA, YUSUKE
THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCE ADDRESSES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-449-731-2
Query Match 3.8%; Score 229.5; DB 4; Length 2843;
Best Local Similarity 19.3%; Pred. No. 1.6e-07;
Matches 266; Conservative 179; Mismatches 510; Indels 421; Gaps 63;
QY 2 HISSEISLSEVAECMGRRO--WKHYDXTCTCSHLNTE---EQQPIA-----LIGS 48
DB 1054 HIIIBEI---KSEQRQSHNQSTTYPVYESTDDKHLKQPHFGQDECVSPKRSRGANGS 1110
QY 49 EDEPQYVMS-SKEISQSNPNHCKTENHRLSEQHNGSQLLLEEDSENNQTSHTSSRTPTP 107
DB 1111 ETNRVGSNHNGINQNSQS---LCQEDDYEDDKPTVYSERYSEEB-----QHEBERPTN 1161
QY 108 GA---TSPTSPPEPPIIDMRPSAKNFCVNGRLITVNAQCKLVAESAATITSSSTNSHH 164
DB 1162 YSIKYNEXRHVDQPTDY-----SLKYATDIPSSQKOSF 1195
QY 165 QHDSQNSGASLPHHISSSSSNNNSNGNRAR---HIAAGARA---TPAATPANSLE 217
DB 1196 SFSKSSSGSGSKTEHMSSESSTSTPSSNAKQONLHPSASGSGQPPKAAATCYVSSIN 1255
QY 218 LKYLITQRAAKWTSMDSMAAQLAOFSLADFNLINSLASQOQOQOQOQOIA-----S 268

DB 1256 QETIQTYVEDTPICFSSRCSLSLSLSAEDICGNQTTQEADSANTLQIAETKGIKIGTRS 1315
QY 269 AVPTTSEVSAALIPALKDTSPSPVDAPLDLSKPSNPSISGCVKSVRACATPTPSGR 328
DB 1316 AEDP-VSEVPVAVSQHPRTK-----SRLOGSSLSSE--SARHKAIVEFPSCA 1358
QY 329 RAYSEEDLSRALQDVVANKLDARKSASQHEORS-ILDNRLLFKMKHD-----Q 376
DB 1359 KSPSK-----SGAQPKSPPEHYQETPLMRSCTSVSLDSFESRSIASSVQ 1406
QY 377 EODHDG-----DELEDSNDDAEAVDSNASTVYVAEFPAKQALRLSHSEHNSDL 428
DB 1407 SEPCSGVSGIISPSDLPS--PGQTMPPSSKTPPPPPOTAQTK-REV---PKGKAPT 1459
QY 429 GEDVGRGSPKGRHPRAGCNASANGAPASIPLDANVLH--TLMAAGIAMPKLDDETQ 486
DB 1460 AEKRESG-PKQ---AAVNAAVQR--VQVLPDADTLHPATSTPTDGFSGSSLSAISL 1511
QY 487 VGDPIKGLLVANSGINNEGLNLTLASQENSNGVASILLQOQOQHQQHQQHQQOQOQ 546
DB 1512 DEPTIQ-----KDELRIIMPVQENDNGN-----ETESQPKESNENQEKEX 1555
QY 547 HVAAYRHRLPKSE-----TPETNSGLDPNDA--SEDPIL 578
DB 1556 TIDSEKDLDDSDDDIEILECIIISAMPTKSRKGRKPAOTASKLPPVARKPQLPQV 1615
QY 579 K-IPFXKSGPSSSSLSRGC-----LVGGHHHPLNNSLSISN-----NSNHSN 624
DB 1616 KLPSPQNLQPKHVSFTPGDMPRVYVEG--TPINSTATSLDLTITESPPNLAQE 1673
QY 625 SHRNQSNRSPHSASPMI-----AAVAQGYSA-----GNSLTSSSSSIQ 665
DB 1674 GVRGAQSGEFPEKRTITPTEGRSTDEAQGKTSSTVITPELDONKAEEDDILAEICNSAMP 1733
QY 666 K-----MMASTIQIINEQSGQESLRNGVSDCSNNGSSSLGYCKPJSIVAKIIGT 719
DB 1734 GKSKHPRPVYKIMQVOOASASSAPKPNOLD-----GKKKPTSPVPPIQON 1782
QY 720 DTSRFGASPNLLSQOHSAHHLTHQOQOQOQASQALCKGRPKRGKY-RVYDRSLVEA 778
DB 1783 TEYRTVRKADSKNNLAEKRVPSDNKD--SKKNLKNNSQDFNDKLPNNEDR----- 1833
QY 779 VKAVQRCGM---SVHRAGSYVGVPH-----STLEYKXVERHILMRPKKEPKQPDLY 827
DB 1834 ---VRGSFARSDPHHYTPIGTGYCFGRNSLSLDDDDVDLSR----- 1876
QY 828 GLTGPAKQLQDKLKAGPHGSKLSNALKNQNNQAAAAAATAATPGLKLPFE 887
DB 1877 ---EKAELRKAKENKESBAKVTSHTELNSQOSANKTOAIA-----KQPINR 1920
QY 888 AGPQALSFQPMFMPQTN-----ATNAYGLDPRNRTTEAMRN-POASNHGMLKMSAD 938
DB 1921 GQPKPI-LQKOSTFPOSSKDIIPDRGAATDEKLQNR---AIENPVCFSHNSLSLSLD 1974
QY 939 MVE---NYVDGIIRITLQASEGNGSAAGNGSNGNGHGHGHGHALLDQLVKKPTL 995
DB 1975 IDQENNMENEPKIKT-EPPDSQEP-----SKPQASGTAPKSFH-----VEDTFV 2019
QY 996 PFTNHRN-----NDYAATCSGAS--GESVYKSGSPMG-----N 1026
DB 2020 CFSRNSLSLSLIDBEDDLQECISSAMPKKKPRRLKGDNEKHSFRNMGILGEDTLTD 2079
QY 1027 YADIKR---ERLADS-----GSSSDHEH----- 1047
DB 2080 LKDIQRPDSHGLSPDSSENFPMKAIQEGANSIVSSLHQAAAAACLSROASDSDSITSLK 2139
QY 1048 ---SASHINNNSDLAHNKNKSGCGGCGGCGNGQOTNGNRSRPTSDSETPSSSFX 1101
DB 2140 SGISLGSFPHLTPDEEKPFTSNK-----GRILLKPGKSTLETETKISSES----- 2185
QY 1102 SGENGQGNHKKMMDLNGSSSSSHIKCE-----SEATGHHSPGHHTTS 1145
DB 2186 KGIGKGRKVVYKSL-ITGKVRNSNSEISQMKQPLQANMPISIRGRWTMHIIPGVNRS 2240

RESULT 10
US-09-125-635-12
; Sequence 12, Application US/09125635
; Patent No. 6562589
; GENERAL INFORMATION:
; APPLICANT: THE UNITED STATES OF AMERICA represented by THE SE
; TITLE OF INVENTION: AIB1, A novel steroid receptor co-activator
; FILE REFERENCE: 48944
; CURRENT APPLICATION NUMBER: US/09/125,635
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 60/049,728
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 12
; LENGTH: 1402
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-125-635-12

Query Match 3.7%; Score 223; DB 4; Length 1402;
Best Local Similarity 19.5%; Pred. No. 1.8e-07;
Matches 214; Conservative 121; Mismatches 384; Indels 378; Gaps 46;

QY 39 EQQP-----IAIAGSEDEPS--QYNHSSKEISQSNHCKTEHRLLEQCHNGSOLLSEEDS 92
DB 568 EONPVESSVCOSSKSDHSEKESKESGSEVSTPRGPLESKGRK-----KLLQLLTTC 619
QY 93 ENNQTSHDSRTPTPGATSTPSPPEPIDWRPSAKCNFCVNGRLITVNAQGLVAESAT 152
DB 620 SSDRGHSS-----LNSPLDPV-----CKSSSVSTPSSG---VSSSTS 656
QY 153 ATSSSTSHIHQSDNSASLSLPHHSSSSSSNNNSGNPARITAAASATPAAATP 212
DB 657 GTVSTSTNVH-----GSLQEKHRLHKLQNGSPAEVAKITAEATGDTSTASC 708
QY 213 ANSLEYLKLTPORAKMTSM-----DSMAOLAQFSLADFNILNSLAQOQQOQ 261
DB 709 GEGTRQEOLESKKKNALRYLLDRDDPSDVLAKELQPOASDSDSKLSQSCSTNPS 768
QY 262 QOQOIASAVPTTSEVSAAISPALKDTPSPSVDAPL-DLSEKSPNSISG----- 312
DB 769 GQEKDPKIKTETNEVSG-----DLQDLAILGLTSSDPYNNPTNGHGGAKOQ 818
QY 313 -----DYKSVACATPTPSGRAYSEEDLSRALQDVVANKLDARKASQHH 358
DB 819 MFAGPSSLGRLSPQFVQSVR-----PPYRAVSLDSPVSGSPVKN-----VSAFPL 868
QY 359 EORSLIDNRLLFKMKHHDQEDHDELESDNDAAEVDNSNSTVYPAE-----PAPA 412
DB 869 FKQPLLAG-----NPRMDSQENYGANNGPNNVNVNPTSSFGDWGLANS 913
QY 413 QLRKLSLSEHNGSDIGEDVDGSPKMGHHPACGNASANOCPASIPLDANVILHTLMLA 472
DB 914 RASRMEPLASSPLGTGADYSATLPR-----PAMG-----GSVPLPLRSNRL----- 956
QY 473 AGIGAMPKLDGTQVDFPKGLLVANSGGINNEGILNLISASQENSNGNASILLQOQOHO 532
DB 957 --PGARPSLQOQO-----QOQ-----QOQOQO 976
QY 533 QHHQOHHOQOQOHH-----VAAVRRRLPKSET--PETSISD--PNAS 573
DB 977 QO 1036
QY 574 EDPILKISFKVSGPSSSS--SLSPGLVGGHHHPUNNNNSISIN-----NSN 620
DB 1037 NRPLRLNSLDLGLGPPNAEGSDERALLDQHTLSTNDAGLEIDRALGIPELVNOG 1096
QY 621 HSSNHRNGSNRSRPSASPMALAAV-----AOGGYSAGNSILLTSSSSSIQKMMASNIOR 674
DB 1097 QALBSKQDVFOGQEAAMVMDKALALYGGTYPAQGPPLQGGFLLQOQSPSFFNSMG----- 1151

QY 675 QINEQGOESLRNGNVSDCSSNNCGSSSLGKPKPSISVAKLIIGTDSRFGASPILLSSQ 734
DB 1152 QISQO-GSFLQO-----GMHRPAGLVRRTPKOL----- 1181
QY 735 HHSAMHLTHQO 794
DB 1182 -----RMQLOQROQO 1222
QY 795 YGYPHSTLEYK-VKERNLHPRKGRKRPQDVLGTLGPANKLOLDLKACPHGSKLSN 853
DB 1223 FFNAQMAAQOQRELMSHLLQOQRMAMMSQOPQAFSPPPN----- 1263
QY 854 ALKQNNQAAAAAATAATPNGIKLPLFAGPQOALSFOPMFEPQTNATNAYGLD 913
DB 1264 -----VTSFPMQVLAGASAMP-----QAPQOQFYPAN-----YGMG 1296
QY 914 -----FNRITAMRNPOASNHGMLKMSAODMVN-----VYDGIIRKTLQASENGSAA 962
DB 1297 QPEPAFOR-----GSSPSPAMMSRMPSSQAMVGHQPTPMY-----QPSDMKGWPS 1345
QY 963 GN-GSNGS-----NGNGHGHGHG-----GALLDQLLVKKTPLPFTVHRNNDYAAT 1008
DB 1346 GNLARNGSFPOQCPAPOGNPAAYNMVHNSGSH--LQVAM--TPWPM----- 1390
QY 1009 CSSASGESVKRSGSPMG 1025
DB 1391 -----SGMPWG 1396

RESULT 11
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352725
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107, 035674
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC
 US-07-741-940-7

Query Match 3.7%; Score 221.5; DB 1; Length 2842;
 Best Local Similarity 19.3%; Pred. No. 6.1e-07;
 Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

```

QY 2 HISSYEISLIERVAEECMGRQ---WKHYODKLTCSHLNIE---EQPIA-----IAGS 48
DB 1053 HIIIEDEI---KQSEQRQSRNQSTTTPVYTESTDDKHLKFQPHFGQECVSPYRSRGANGS 1109
QY 49 EDEPSQYVNH-SKEISQSNPNHCKTENHRLQOHNQSOQLLEEDSENNQTSRTPPTP 107
DB 1110 EFNRYGSHHGINQNVQS---LCQEDDYEDDKPTVYSEYSEE-----QHEEBERPTN 1160
QY 108 GA---TSTPSPEPPIIDMRPSAKNFCVNGRLITVNAQGLVAESAATATSTNSHIH 164
DB 1161 YSIKYNEEKRHVDQPIDY-----SLKYATDIPSSQKQSF 1194
QY 165 QHDSQSNSSASLPHHSSSSSSNNNSGGRAR---HIAASARA---TPAATATANSIE 217
DB 1195 SFSKSSSQSGSKTEHMSSESTSTPSSAKRONOLHPSSAQSRGQPOKAATCKVSSIN 1254
QY 218 LYKLITOPAAKWTSMDSMAAQLAOFSLADPNLINSLSAQOQQOQQOQIA-----S 268
DB 1255 QETITCYCEDTPICFSSSSLSLSAIDEICGCTOTQOASANTLOIAETKEKIGTRS 1314
QY 269 AVPTTSEVSAASIPALKDTPSPVDAEPLDLSKSPNSISGDKVYACATPTPSGR 328
DB 1315 AEDP-VSEVPVAVSQHPRTK-----SSRLQSSLSISE--SARHKAVEFSSGA 1357
QY 329 RAYSEDELSRALQDVANKLDARKASQHEQRS--ILDNRLLFKMHHD-----Q 376
DB 1358 KSPSK-----SGAQTPKSPPEHYVQETPLMFSRCTSVSLSLSEFESRSIASSVQ 1405
QY 377 EODHDC-----DELESDNDAAEVDNSASTVYPAFAOLRLKLSHEHNSDL 428
DB 1406 SEPCCGMVSGIISPDLPDS--PGOTMPSPRSKTPPPPTQATQK-REV---PKNKAPT 1458
QY 429 GEDVDGRSPKQGRHDPACGNASANOGAPASIPLDANVLH--TLMLAAGIAMPKLDQTQ 486
DB 1459 AEKRESG-PKQ---AAVVAAYQR---VQVLPDADTLHFAFATESTPDGSSCSLSLSL 1510
QY 487 VQDFIKGLLVANSGIMNEGILNLSASQENSGNASLLQQOQHQHQHQHQOQQOQ 546
DB 1511 DEPFIO-----KQVELRLIMPVQENDNGN-----ETESQPKESNDEKEK 1554
QY 547 HVAAYRHLRPKE-----TPETNSSLDPPDA---SEDPIL 578
DB 1555 TIDSEKDLDDSDDDILEECITISAMPTKSRKAKKPAQASIKLPPVAKPQQLPVY 1614
QY 579 K-IPSFKVGSPASSSSSLSPG-----LVGHHHPLNNSNLSISN-----NSNHSN 624
DB 1615 KLPFQONRLQPOKHVSTFGDDMPRVYCVEG--TPINSTATSLDLTIESPNEIAGE 1672
QY 625 SHRNSNSNPSHASPM-----AAVAQCGYSA-----GNSLITSSSSSIQ 665
DB 1673 GVRGAQSGEPEKDTIPTEGSTDEAOGKTSVTIPELDNKAEBGGIILAECCINSAMP 1732
QY 666 K-----WMAANIQRQINQSGQESLRNGNVSDCSSNNGSSSLYKPKPSIVAKITGT 719
DB 1733 KQKSHKPRVKKIMQVQQAASSAPKNGQD-----GKKKKPISPVKPIPON 1781
QY 720 DTSRFGASPNLISQOHSANHLTHOQOQOQLSAQBALGKTRPKRGKY-RNYDRDSLVEA 778

```

```

DB 1782 TEYRTRVRKNAKDSKNLNAERVFSDNKD---SKQNLKNNSKDFNDKLPNNEDR----- 1832
QY 779 VKAVORGEM---SVRAGSYGVPH-----STLEKVKVERHLMPRKREPKQPDLY 827
DB 1833 ---VRGSFAPDSPHHYTPIGTPTCYFSRNDLSLADPDDDVDLSR----- 1875
QY 828 GLTGPAKLOLDKLCAGPHGSKLSNALKNQNNQAAAAAATAATPGLKLPFE 887
DB 1876 -----EKAELRKAKENKESKAVTSHTELTSNQSAAKTQALA-----KQPINR 1919
QY 888 AGPQALSTQPMNFWPQTN-----ATNAYGLDFNRTITAMRN-POASNHHGLKMSAD 938
DB 1920 GQPKPI-LOKOSTFPQSSKDIIDRGAATDEKLONF-----AIENTPVCFSHNSLSLSD 1973
QY 939 MVE---NYVDGIRKTLQASGNGSAANGNSGNGCHGHCHALLDOLVYKTPPL 995
DB 1974 IDQENNNKENEPIKET-BPPSQGEP-----SKPQASGVAPKSFH-----VEDPVP 2018
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVYKRSQSPMG-----N 1026
DB 2019 CFSRNSLSLSLIDSEDDLQECISSAMPKKKPKRLKGDNEKHSPRMNGILGEDLTLD 2078
QY 1027 YADIKR---ERLADS-----GSSDDEH----- 1047
DB 2079 LKDIORPDSEHGLSPDSSENFPMKAIQEGANSIVSLHQAIAAACLSROASSPSDILSLK 2138
QY 1048 ---SASHINNNSDLAHKKNKSGGGGGGNGQNTGNRSSMTSRDSETPASSRFK 1101
DB 2139 SGISLGSFPHLTPDOEKRPFTSNK-----GPRILKPEKSTLETKKIESES----- 2184
QY 1102 SCENGQOQNHKMDLNGSSSSSHIKCE-----SEAAIGHNSPGHHTTS 1145
DB 2185 KGIGKGKVVYKSL-ITGKVRNSNSISGOMKQPLQANMPSISRGRTMIIHPVRNNS 2239

```

RESULT 12

US-08-289-548A-7
 ; Sequence 7, Application us/08289548A

Patent No. 5648212

GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allgretti, LTD

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/289,548A

FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107,46943

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match      3.7%; Score 221.5; DB 1; Length 2842;
Best Local Similarity 19.3%; Pred. No. 6,1e-07;
Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSEYELREVAECMGRRO--WKHYODKUTCSHLINE--EQQPIA-----IAGS 48
DB 1053 HIEDEI--KQSEORQSRNOSTTYPVYTESTDHLKFOHPHGOQECVSPYRSRGANGS 1109
QY 49 EDEBSQVYNS-SKEISQGNPNHCKTENHRLSEQHNGSOLLEBEDSENNQTSHDSRTPTP 107
DB 1110 ETNRVSGHGINQNVQS--LCQEDYEDDKPTMYSERYSFE-----QHEBERRTN 1160
QY 108 GA--TSTPPPPPEPIDMRPSAKNFCVNGRLITVNAQGLVAESAATATSSSTNSHIH 164
DB 1161 YSIKYNEEKRHVDOPIDY-----SLKATATDIPSSQXQSF 1194
QY 165 QHDSNNSASLPHHISSSSSNNNSGNRRAR--HTAASARA--TPPAATANGLE 217
DB 1195 SPSYSSSQSSSTHEBSSSSSESTSTPSSNAKQONQHLPSAQRSGQPOKATCKVSSIN 1254
QY 218 LYKLTQRAAKTMSDMAAQLAQLADLADFNILNSLASQOQOQOQOQOIA-----S 268
DB 1255 QETIQTVEEDTPICFGRCSLSLSADELICQNTOTQEDASANTLQIAETKEKIGTRS 1314
QY 269 AVPTTSVSAASIPALKDTPSPSVDAPLDLSKPSNNSISGDKSVRAQATPTPSGR 328
DB 1315 AEDP-VSEVPVAVSQHPRTK-----SSRLQSSSISSE--SARHKAVERFSSGA 1357
QY 329 RAYSEEDLSRALQDVVANKLDARKSAQHEQRS--ILNRLFKMHHND-----Q 376
DB 1358 KSPSK-----SGAOTPKSPPHVVOETLHFSRCTSVSLDSFESRSIASSVQ 1405
QY 377 EQDHG-----DELESDNDAAEAVDSNASTPVPAEFARQRLKLSLSEHNGSGL 428
DB 1406 SEPCSGWVGIIISPDLPS--PGQTPSPRSKSTPPPPQTAQTK-REV---PKNKAPT 1458
QY 429 GEDVDKRGPKMGRHACGNASANGAPASIPLDANVLH--TLMAAGIGAMPKIDETQT 486
DB 1459 AEKESG-PKQ--AAVNAAVOR--VQVLPDADTLHFAFESTPDPFSSSSLSLSL 1510
QY 487 VGDPIKGLLVANSGIMNEGILNLISASQENSNGASLLQQOQOQHQQHQHQOQOQO 546
DB 1511 DEPIQ-----KVELRLMPVOENDNGN--ETESQPRESENDEKAEK 1554
QY 547 HVAAYRHRLPKSE-----TPETNSLIDPNDN--SEDPIL 578
DB 1555 TIDSEKDLDDSDDDDIILEECISAMPTKSSRAKAPQATASLPPVARKPSQLP 1614
QY 579 K-IPGFVSGPAASSSLSPGG-----LVGHHHMLNNNSLSISN-----NSNHSN 624
DB 1615 KLTPSQNLQPOKIVSFTPGDDMPRVYCVEG--TFINSTATSLDLTIESPVELAGE 1672
QY 625 SHRNSNNSPHSASML-----AAVAAGGISA-----GNSLITSSSSSIQ 665
DB 1673 GVRGAQSGEPEKRDITPTEGRSTDEAGGKTSVTIPELDNKAEBEDDIAECINSAMP 1732
QY 666 K-----MMAAGNIQROINQSGQESLRNGNVSDCSNNGSSSLGKPKPSIAKIGCT 719

```

```

DB 1733 KKGSHKPRVKKINDQVQASASSAPNKQLD-----GKXKPSVPKPIQN 1781
QY 720 DTSRFGASPNLLSQOHSAAHLLTHQOQOQOQISAOEALGKTRPKRGKY-RNYDSDSLVEA 778
DB 1782 TEYTRVAKNADSKNNNAERVFSDNKO--SKQONKNNSKDPNDKLPNNEDR----- 1832
QY 779 VKAVORGEM--SVHRAGSYGVPH-----STLEKXVERHMLMPRKKEPRQPDLV 827
DB 1833 ---VRGSFARSDPHHTPIGTPYCFSRNDSLSLDPDDDDVDLSR----- 1875
QY 828 GLTGPANKLQDLKAGPHGSKLSNALKNQNNQAAAAAATAATNGIKLPLFE 887
DB 1876 -----EKAEELRKAKENKESAKVTSHTELTSNQASANKTOAIA-----KQPINR 1919
QY 888 AGPOALSFQPMFMPQTN-----ATNAYGLDPNRTITEMMRN-POASHHGLMKSD 938
DB 1920 QQPKPI-LQKSTPQSSKQIDPRGAATDEKIQNF-----AIENTPVCFSHNSLSLSLD 1973
QY 939 MVE--NYVDGIIKRTLOASEGNGSAAAGNSGNSGNGHGHGHALDLQLLVKKTPL 995
DB 1974 IDQENNNKENPIKET-FPPDSQGP-----SKPOASGTAAPKSFH-----VEDTPV 2018
QY 996 PFTNHRN-----NDYAATCSSAS-----GESVKRSGSPMG-----N 1026
DB 2019 CFSRNSLSLSISSEDDLLQECISAMPKKKKRSLKQDNKESPRNMGGILGEDTLTD 2078
QY 1027 YADIKR--ERLSADS-----GGSDDEH----- 1047
DB 2079 LKDIORPDEHGLSPDSENPFWKAIQEGANSIVSLQAAAAACLSRQASDSDSISLSLK 2138
QY 1048 ---SASHNNNSDLDANKNKSAGGGGGGQNGQNTNGNRSRMTSRDSETDASSFK 1101
DB 2139 SGILGSPFHLTPQOEKPTSNK-----GRILLKPEKSTLETYKLESSE----- 2184
QY 1102 SGENGQGNHMKMDLINGSSSSSHIKCE-----SEATGHHSPGHTTTS 1145
DB 2185 KGIGKQKVVYKSL-ITGKVRNSNSELISGQMQLQANMPSISRGTMHIEVRRNS 2239

RESULT 13
US-08-452-654-7
Sequence 7, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995

```

CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 3.7%; Score 221.5; DB 1; Length 2842;
Best Local Similarity 19.3%; Pred. No. 6.1e-07;
Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSEISLIERVAEEMGR--WKHYODKLTCSHLNTE---EQPIA-----IAGS 48
DB 1053 HIIIEDEI---KQSEQRQSRNQSTTPVYTESTDDKHLKQPHFGQCEVSPYRSGANGS 1109
QY 49 EEPSPQVNHNS--KEKISQSNPNHCKTENHRLBEOHNGSOLLEBEDSNNNTSDSRTPTP 107
DB 1110 ETRNRYGSHNGIOWNSQ--LCQEDDYDDKPTNTSERYSEB-----QHEEERPTN 1160
QY 108 GA---TSTPTPEPPIDMPRPSPAKCNFCVNGRLTVNAQGLVAESAATATSSSTGSHIH 164
DB 1161 YSIKYNEKRAHVDPIDY-----SLKYVTIDIPSSQKQSF 1194
QY 165 QHDSQNSNSASLPHHISSSSSSSNNNSGGRAR---HIAASARA---TPAATPANSLIE 217
DB 1195 SFSKSSSGQSSKTEHMSSESTSTPSSNAKRONQLHPSSAQSRSQPOKAAATCKVSSIN 1254
QY 218 LYKLTQRAKMTSMDSMAQLAOPSLADPFLNLSLAQQQQQQQQQQA-----S 268
DB 1255 QETIQYCVEDTPTICFSCSSLSUSSADEIGCNOTTOEADSAANTLQIAIEIKKIGTRS 1314
QY 269 AVPTTTEVSAASISPALKOTPSPSVDAPLDSKPSPNSSISGVKSVACATPTPSGR 328
DB 1315 AEDP--VSEVPVAVSQHPRTK-----SSRLQSSSLSS--SARHKAVERSSGA 1357
QY 329 RAYSEBDLSRALQDVVANKLARKSAQHNEQPS--ILDNRLFRMKHND-----Q 376
DB 1358 KSPSK-----SGAQTTPKSPREHYVQETPLMFSCRCTSVSLDSFESRSIASSVQ 1405
QY 377 EGDHNG-----DELEDSNDAAEVDNSASTPYPAEFARAQRLKSHSEHNGSL 428
DB 1406 SPPCGMWGSIISPSDLPDS--PGQTMPRSRKTTPRPOTAQTK--REV---PKKAKT 1458
QY 429 GEDVDRGSPKGRHPRACNANOGAPASIPLDANVLH--TLMLAAGIGAMPKLDDETQ 486
DB 1459 AKRESG-PKQ---AAVVAAYOR---VQVLPDADLTLHFAESTPDGSCSSSLASL 1510
QY 487 VQDFIKGLIVANSGLIMNEGILLNLASQENSGNANSLILQOOQHQQHQQHQQQQQ 546
DB 1511 DEPTIQ-----KQVELRIMPVQENDNGN-----ETSEQPKSENQEKEMAK 1554
QY 547 HVAAYRHRLPKSE-----TPTNSSLDPDA---SEDPIL 578
DB 1555 TIDSEKDLLDDDDDDIELEECIISAMPTKSSRAKKAPPAQATSKLPPVAAKPPQQLPVY 1614
QY 579 K-IPSEKVGSPASSSSSLSPG-----LVGGHHHPLANNNSLSISN-----NSNHSN 624

DB 1615 KLLPSONRLQPOKHSFTPGDDMPRVYCEG--TFINSTATSLDLTIESPPNELAGE 1672
QY 625 SHRNSNNSPHSASML-----AAVAAGQISA-----GNSLITSSSSSIQ 665
DB 1673 GVRGAQSGEEPEKRDITTEGRSTDEAGGKTSTYTIPELDNKAEEBDIIEACINSAMP 1732
QY 666 K-----NMAANIQRQINEQSGEELRNGVNSDCSNNNGSSSLGKPPSISVAKITGOT 719
DB 1733 KKKSHKPPRVKIMQVQAASASSAPKNOLD-----GKKKKPSPVPIQON 1781
QY 720 DTSRFGASPNLLSQOHSAAHLTHQOQQOQLSQAQALCKGRPKPKGY--RNYDRSLVEA 778
DB 1782 TEYTRVRKADNSKNNLNAERVFSDNKO--SKQNLKNNSKDFNDKLPNNEDR----- 1832
QY 779 VKAVRGEM---SVIRAGSYGVPH-----STLEKVKERHLMRKRKEPKQPDLV 827
DB 1833 ---VRGSFAPDSPHHYTPTEGTPTFCFSRNDLSLSLDFDDDDVDLSR----- 1875
QY 828 GLTPANKLOLDKLAGPHGSKLSNALKNONNQAASAAAAAATAATPGLKPLPE 887
DB 1876 -----EKAELRKAKENESKAKVTSHTELISNQAASAKTKQAIA-----KQPIRK 1919
QY 888 AGPQALSFQPMFWPQTN-----ATNAYGLDENRITTEAMRN--POASNHRLMKSAQD 938
DB 1920 GQPKPI-LQKSTFPQSKDIPDGAATDEKLQNF-----AIENTPVCFSHNSLSLSLD 1973
QY 939 MVE---NYYDGIKRTLOASEGNSAAGNSGNSGNGHGHGHGHALLDOLLYKPTPL 995
DB 1974 IDQENNNKENEPIKET--EPPSQGEP-----SKPOASGYAPKSFH-----VEDPTP 2018
QY 996 PPTNHRN-----NDYAATCSAS-----GESYKRSQSPMG-----N 1026
DB 2019 CFSRNSLSLSISDEBDLLOECISSAMPKKKKRBLKGDNEKISPRNKGGLGDLTLD 2078
QY 1027 YADIRK---ERLSADS-----GSSSDEEH----- 1047
DB 2079 LKDIQRPSEHGLSPDSENFMWKATIQEGANSIVSLHQAAAAACLSRQASSDSISLSLK 2138
QY 1048 -----SASHINNNSDLAHHKNKSGGGGGGCGGQCTNGNGRSSMTSPDSSETAASERK 1101
DB 2139 SGISLSPFHLTPDDEEKPFTSNK-----GPRILKPEKSTLETETKIEBS----- 2184
QY 1102 SGENGGQONHKMDLNGSSSSSHIKCE-----SEATGHHSPGHHTS 1145
DB 2185 KQIKGKGVYKSL-ITGKVRNSSEISGQMKOPLQANPSPISRGRTMHI PGVRNSS 2239

RESULT 14
US-08-449-731-7
Sequence 7, Application US/08449731
Patent No. 6413727
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
ANAND, RAKESH
CARLSON, MARY
GRODEN, JOANNA
HEDGE, PHILIP J.
JOSLYN, GEOFF
KINZLER, KENNETH
MARKHAM, ALEXANDER F.
NAKAMURA, YUSUKE
THALIYERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7

Query Match 3.7%; Score 221.5; DB 4; Length 2842;
Best Local Similarity 19.3%; Pred. No. 6.1e-07;
Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSEILSERVAEECMGRQ---WKHYODKLTCSHLNTE---EQPIA-----IAGS 48
DB 1053 HIIIEI---KSEBORQSRNSTYPTVTESTDDHLKQPHFGQECVSPRSGANGS 1109
QY 49 EDEPSQVHNS-SKEISGNPNHCKTENHRLQOHNGSQLLEEEDENNQTSRPTP 107
DB 1110 ETRNVRSGNHGINONVQS---LQEDDYEDDKPTVYSRYSBEE---QHEEERPTN 1160
QY 108 GA---TSTPSPPEPIIDWRPSAKCNFCVNGRLTYNAGQKVAEBAATATSSSTNSIH 164
DB 1161 YSIKYNEEKRHVDQIDY-----SLKYATDIPSSQKQSF 1194
QY 165 QHSDSDNSGASLPHHIISSSSSSNNNSGNRAR---HIAASARA---TPAATPANLE 217
DB 1195 SFSKSSSSGQSKTEHMSSESTSTPSSNAKQONLHSSAOSRSGQOKAATCVSSIN 1254
QY 218 LYKLITQRAAKKTSMDMAAQLQPSLADPNLINSLSAQOQOQOQOIA-----S 268
DB 1255 QETITQYCEVEDPTLICFSRCSSLSLSAEDETGICQTTQEDASANTLQIAETIKETGRS 1314
QY 269 AVPTPTSEVSAALSPALKDTPSPSPVDAPLDLSSKSPSSNISGDKVSRACATPTPSGR 328
DB 1315 AEDP-VSEVPVAVSQHPTK-----SSRLQSSSSISE--SARHKAVERFSSGA 1357
QY 329 RAYSEEDLSRALQDVANKLDARKASQHEQRS-ILDNRLFKMGHD-----Q 376
DB 1358 KSPSK-----SGAQTPKSPRPHYQETPLMFSRCTSVSLSDFESRSTASSVQ 1405
QY 377 EQDHG-----DELEDSDDAEVDSNASTFVPAEFAPARQLRLKLSHSENGDL 428
DB 1406 SEPSCGWSGIISPDLPDS--PGQTWPPSRSKTPPPPPQATQK-REV---PKKAPPT 1458
QY 429 GEDVDKSGPRKGRHACGNASANOAPASIPLDANVLH--TLMAGIGAMPKIDETQT 486
DB 1459 AEKESG-PKQ---AAVNAAVOR---VOVLDADDTLLHFATESPTDPSGSSSLALSL 1510

QY 487 VGDFIKGLLVANSGGIMNEGILNLISAOENSGNASILLQOQOHOQHQQOQOQOQ 546
DB 1511 DEPIQ-----KDVLRITMPVQENDNGN-----ETSEBQPRESENQEKEMK 1554
QY 547 HVAAYRHRPKSE-----TPTNSSLDPNDA---SEDPIL 578
DB 1555 TIDSEKDLDDSDDDIEILECTISAMPTKSRKAKKPAQTASKLPVPARKSQULPVY 1614
QY 579 K-IFSKVSGPSSSSSPG-----LVGHHHPPLNNNSLSISN-----NSNHSN 624
DB 1615 KLLPSQNLQPKQKVSFTPGDMPRVYCVES--TPINSTATSLDITLTPSPNELAAGE 1672
QY 625 SHRNGSNRSPHSASPMU-----AAVAQGYSA-----GNSLITSSSSSIQ 665
DB 1673 GVRGAQSGEPEKRDITPTBGRSTDEAGGKTSSVTIPELDONKABESDILAECINSMP 1722
QY 666 K-----MMASNIRQINEQSGESLNGVNSDCSSNNGSSSIQYKPSISVAKIIGCT 719
DB 1723 KKGSHKPPRVKINDVOQASASSASPKNQLD-----GKKKKPSPVYKPIPON 1781
QY 720 DTSRFGASPNLLSQOHSANHLTHQOQOQQLSAQBALGKGRPKRGY-RVYDRDSLVEA 778
DB 1782 TBYTRVRKADSKNNLAEVRFSDNKO---SKQNLKNSKDFNDKLPNNEDR----- 1832
QY 779 VKAVQSGEM---SVHRAGSYGVPH-----STLEKVKERHLMRPKKEPKPQDLY 827
DB 1833 ---VRGSFAPDSPHHTPIEGTPYCFSRNDSLSLDPDDDDVLSR----- 1875
QY 828 GLTGPAKQLQDKLACPHGSGKLSNALKNONNAAAAAATAATNGIKLPLFE 887
DB 1876 ---EKAEELRKAKENKSEAKVTSHTELSNOSANKTOAIA-----KQPINR 1919
QY 888 AGPOLSPQPMFPMPTN-----ATNAYGLDFNRTTEAMR-POASNHGMLKMSQD 938
DB 1920 GQPKPI-LQKSTPPOSKDIPDGAATDEKQNF---AIENTPVCFSNSSLISD 1973
QY 939 MVE---NVYDGIIRKTLQASEGNSAANGSNGNGHGHGHGHALLDQLVKTPL 995
DB 1974 IDQENNNKENEPIKET-EPPSQGEP-----SKPQASGVAPRSPH-----VEDPIV 2018
QY 996 PFTNHRN-----NDYATCGSAS-----GESVKRSGSPMG-----N 1026
DB 2019 CFSRNSSLSSLSISEDDLQECISAMPKKKPSRLKGDNEKHSPPNMGGILGEDTLTD 2078
QY 1027 YADIKR---ERLADS-----GGSSDEH----- 1047
DB 2079 LKDIQRPDSHEGLPDSBNFDWKAIQGANSYVSLHQAALACLSROASSDSISLSK 2138
QY 1048 ---SASHINNNSDLAHNKNKSGGGGGGNGGQTNNGRSSRMTSRDDSETDAS9FK 1101
DB 2139 SGISLSPFHLTPQEEKPFTSNK-----GPRILKPGKSLFTKXIESES----- 2184
QY 1102 SGENGQOQNHQMDLNGSSSSSHIKCE-----SEATGHSPGHHTTS 1145
DB 2185 KGIGGKKVYKSL-ITGKVRNSSEISGQKOPLOANPSPISRGRTHHIPGRNNS 2239

RESULT 15
US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

```

: TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
: NUMBER OF SEQUENCES: 102
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Witcoff, Ltd.
: STREET: 1001 G Street, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20001-4598
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/452,655B
: FILING DATE: 25-MAY-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,548
: FILING DATE: 12-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/741,940
: FILING DATE: 08-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 1107.49964
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2843 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-452-655B-2

Query Match      3.7%; Score 221.5; DB 1; Length 2843;
Best Local Similarity 19.3%; Pred. No. 6,1e-07;
Matches 265; Conservative 179; Mismatches 511; Indels 421; Gaps 63;

QY 2 HISSEIELSERVAEECEWGRRO---WKHYQDKTSGHLINE---EQQPIA-----LNGS 48
DB 1064 HIEIEI---KQSEOROSNOSTTYPVTESTDDKHLKQPHFGQECVSPYRSGANGS 1110
QY 49 EDEPSQVYNS-SKEISQSNPNHCKTENHLEQOHNGSOLLEEDSEENNQTSHTSSRTPTP 107
DB 1111 ETNRVSGNHGINQNVNSQS---LQCEDYEDDKPTVYSERYSEE-----QHEEERTTN 1161
QY 108 GA---TSTPSPPPEPIDMRPSAKNFCVANGRLLTVNAQGLVAESAATATSSSTNSHIH 164
DB 1162 YSIKYNEEKRAHVDPIDY-----SLKYATDTPSSQKQSF 1195
QY 165 QHDSNNSASLPHHSSSSSSNNNSGNRAR---HIAASARA---TPAATFANLE 217
DB 1196 SPSKSSSSQSSKTEHMSSESTSTPSSNAKRONOLHPSAQSRSQPOKATCKVSSIN 1255
QY 218 LYKLTORAAKWTSMDSMAAQLQFSLADFNLLINSLASQOQOQOQOQOIA-----S 268
DB 1256 QETIGTCEVDPTPICFSRGSSSLSSLSADELICGQTTQOEAQNTLQIAETKEKIGTRS 1315
QY 269 AVPTTTSVSAALISFALKDTPSPVDAPLDISKSPNSGISGVKSVACATPTPGR 328
DB 1316 AEDP-VSEVPVAVSQHPRTK-----SSRLQSSLSSE--SARHKAVEFFSSGA 1358
QY 329 RAYSEEDLSRALQDVAVAKLARKSASQHEQRS--ILDNRLLPKMKHND-----Q 376
DB 1359 KSPSK-----SGAQTPKSPPEHYVQETPLMFSRCTSVSSLSDFESRSIASSVQ 1406
QY 377 EODHDG-----DELEDSNDAAEAEVDSNASTPVYPAFAQLRLKLSLSEHNGSDL 428

```

```

DB 1407 SEPSCGVNSGIISIDLPDS--PGQTMPPSKTKTPPPQTAQTK-REV----PKNKAPT 1459
QY 429 GEDVDRGS PKMGRHPACNAGANOGAPASIFLDANVLH--TLMLAGIAMPKLDETQT 486
DB 1460 AKEESG-PKO---AANAAVQR---VOVLPDQTLHFPTSTPDGFCSSSSISALS 1511
QY 487 VGDPIKGLLVANSGGINNEGILLNLSSAQENSNGNASLLLOQOOHQOHQOHOQOQOQO 546
DB 1512 DEPIQ-----KDELRLIMPVQENDNG--ETESQPKESNENQEKAEK 1555
QY 547 HVAAYRHRPKSE-----TPETNSLDPNDA--SEPI 578
DB 1556 TIDSEKDLDDSDDDIELEECIISAMPTKSRKAKPAQTASKLPPVARKSQQLPV 1615
QY 579 K-IPSFVXSGPSSSSISPGG-----LVGHHHPHNNNSLSTN-----NSNHSN 624
DB 1616 KLBPQNRLOPKHVSFTPGDMPRVVCVEG--TPINSTATSLSLDTLTPSPNELAAGE 1673
QY 625 SHRNGNRS PHSAS PML-----AAVAQGYSA-----GNSLLTSSSSSIQ 665
DB 1674 GVRGAQSGEFKEDTITPBGSTIDEAQGKTSVTIPELDONKAEEDDILAECINSAMP 1733
QY 666 K-----MNASNIORQINEQSGESLRGNVSDCSSNNGSSSLGKYPKPSISVAKITGT 719
DB 1734 KAKSHKPPRVKKIMDOVOQASASSAPYKQOLD-----GKKKKPTSPVPIPN 1782
QY 720 DTSRFGASPNLSQOHSANHLTHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 778
DB 1783 TEYRTRVRKNAKDSKNLAEVVFSDNKO--SKQNLKNNKDPFNKPNNEDE-- 1833
QY 779 VKAVORGE--SVYRAGSYGVPH-----STLEKVKERHLMRPREKPKQPOLV 827
DB 1834 ---VRGSFARSDPHHYPIGTPYCFGRNDSLSLDDDDDDVLSR----- 1876
QY 828 GLTGPAKQLQDKLACGPHGSKLSNALKNONQAAAAAATAAATNGCLKPLFE 887
DB 1877 ---EKAEILRKAKENKESAKVTSHETLSNOQSANKTQAI-----KQPINR 1920
QY 888 AGPQALSFQPMFMFQTN-----ATNAGGLDFNRTTEAMRN--PQASNHGMLKMSQD 938
DB 1921 GQPKPI-LQKOSTPQSSKDIIDRGAATDEKLONF--AIETPVCFSHNSLSLSLD 1974
QY 939 MVE---NVYDGLIRKTLQASEGNSAAGNSGNSGNGHGHGHALLDQLLYKTKPL 995
DB 1975 IDQENNKENPIKET--EPPOSQEP-----SKPQASGYAPKSFH-----VEDTPV 2019
QY 996 PTNHRN-----NDYAATCSAS-----GESVKRSGSPMG-----N 1026
DB 2020 CFSRNSLSLSISIDSEDLLECISAMPKRRKSRLLKGNDEKISPRMGGILGEDTLTD 2079
QY 1027 YADIKR---ERLSADS-----GSSSDEEH----- 1047
DB 2080 LKDIORPSEHGLSPDSENFPMKAIQEGANSIVSLHQAATAACLSROASSDSISLSLK 2139
QY 1048 ---SASHINNNSDLAHNKNSGGGGGGGNGQNTNGNRSMTSRDSETPASSFX 1101
DB 2140 SCISLGSFHLTPPOEEKPFTSNK-----GRILKPGKSTLETLEKISSES----- 2185
QY 1102 SGENGGQONHKMDLNGSSSSSHIKCE-----SEAAIGHSPGHHTTS 1145
DB 2186 KGIRGKKVYKSL-ITGKVRNSNLSISGQMKPLQANMPSIRGRMTMHIPIVRNNS 2240

```

Search completed: October 29, 2003, 12:32:08
Job time : 45 secs

THIS PAGE BLANK (USP)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Gen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:21:56 ; Search time 28 Seconds

(without alignments)
1956.647 Million cell updates/sec

Title: US-10-016-768a-10

Perfect score: 6030

Sequence: 1 MHSSYEISLEKVAEECMGR.....ILHEKLAQIKAEQVDADQL 1165

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	311.5	5.2	2038	1 FSH_DROME	P13709 drosophila
2	299.5	5.0	2175	1 HMCU_DROME	P10180 drosophila
3	285.5	4.7	1211	1 BUN2_DROME	Q24523 drosophila
4	272.5	4.5	1073	1 HR38_DROME	P49689 drosophila
5	272.5	4.5	1669	1 ASX_DROME	Q9V727 drosophila
6	271	4.5	1365	1 SUZ2_DROME	P25172 drosophila
7	259	4.3	1028	1 OVO_DROME	P51521 drosophila
8	257.5	4.3	1077	1 HLES_DROME	Q02308 drosophila
9	249.5	4.1	1556	1 PROS_DROVI	Q9U641 drosophila
10	249	4.1	1165	1 YNF4_YEAST	P53950 saccharomyc
11	246.5	4.1	880	1 BRC4_DROME	Q24206 drosophila
12	244.5	4.1	883	1 E74B_DROME	P11536 drosophila
13	243	4.0	1507	1 SIVA_DROME	Q24167 drosophila
14	238.5	4.0	1403	1 PROS_DROME	P32870 drosophila
15	237	3.9	2248	1 CYA1_DROME	P55198 homo sapien
16	231	3.8	1093	1 AP17_HUMAN	P17672 drosophila
17	231	3.8	1394	1 E75B_DROME	P45447 drosophila
18	230.5	3.8	865	1 E75A_DROME	P28166 drosophila
19	228	3.8	1060	1 ZFH1_DROME	Q9EPQ8 mus musculu
20	225.5	3.7	1983	1 TF20_MOUSE	Q01295 drosophila
21	223	3.7	727	1 BRC1_DROME	Q9U640 homo sapien
22	223	3.7	1960	1 TF20_HUMAN	Q9U640 homo sapien
23	222	3.7	1253	1 DSRP_HUMAN	P25054 homo sapien
24	221.5	3.7	2843	1 APC_HUMAN	Q92212 candida alb
25	221	3.7	1230	1 ST20_CANAL	P14196 dictyostell
26	220.5	3.7	448	1 AAC2_DICDI	P54637 dictyostell
27	220	3.6	960	1 YMX6_YEAST	Q04893 saccharomyc
28	220	3.6	989	1 PTP3_DICDI	P10105 drosophila
29	219.5	3.6	1140	1 YMX6_YEAST	Q9U105 drosophila
30	216.5	3.6	635	1 HMLA_DROME	Q03825 saccharomyc
31	216.5	3.6	1403	1 CTC_DROME	P21519 drosophila
32	215	3.6	758	1 YMX8_YEAST	
33	214	3.5	1596	1 MAM_DROME	

34	212.5	3.5	1317	1 GAP_CAEEL	P34288 caenorhabdi
35	211.5	3.5	1237	1 E75A_DROME	P17671 drosophila
36	211	3.5	918	1 YMX8_CAEEL	P34487 caenorhabdi
37	211	3.5	1443	1 E75C_DROME	P13055 drosophila
38	210.5	3.5	1780	1 YK26_CAEEL	P34333 caenorhabdi
39	209	3.5	1043	1 APC_MOUSE	P33244 drosophila
40	209	3.5	1845	1 TF20_MOUSE	Q61315 mus musculu
41	208.5	3.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
42	208	3.4	1597	1 SOL_DROME	P27398 drosophila
43	206.5	3.4	931	1 LUG_DROME	Q9FUY2 arabidopsis
44	206	3.4	1319	1 MNI_HUMAN	Q10571 homo sapien
45	205	3.4	805	1 E2F_DROME	Q27368 drosophila

ALIGNMENTS

RESULT 1
FSH_DROME
ID P13709: P13710: STANDARD: PRT: 2038 AA.
AC 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Female sterile homeotic protein (Fragile-chorion membrane protein).
GN FS(1)H OR FSH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89276730; PubMed=2567251;
RA Haynes S.R., Mozer B.A., Bhatia-Dey N., David I.B.;
RT "The Drosophila fish locus, a maternal effect homeotic gene, encodes
RT apparent membrane proteins."
RL Dev. Biol. 134:246-257(1989).
CC -!- FUNCTION: REQUIRED MATERNALLY FOR PROPER EXPRESSION OF OTHER
CC HOMOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC -!- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC -!- SIMILARITY: Contains 2 bromodomains.
CC -!- SIMILARITY: Contains 1 ET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M23221; AAA28540.1; -
CC EMBL: M23222; AAA28541.1; ALT_TERM.
CC EMBL: M15762; AAA70424.1; -
CC EMBL: M15763; AAA70423.1; -
CC EMBL: M15764; AAA70422.1; -
CC FIR: A43742; A43742.
CC HSP: Q92831; I891.
CC FlyBase: FBgn004656; Fg(1)h.
CC InterPro: IPR001487; Bromodomain.
CC Pfam: PF00439; bromodomain.2.
CC PRINTS: PR00503; BROMODOMAIN.
CC SMART: SM00297; BROMO.2.
CC PROSITE: PS00633; BROMODOMAIN_1; 2.
CC PROSITE: PS50014; BROMODOMAIN_2; 2.
CC Developmental protein: Bromodomain; Transmembrane; Repeat.
FT DOMAIN 51 123 BROMODOMAIN 1.
FT DOMAIN 495 567 BROMODOMAIN 2.
FT DOMAIN 945 1106 ET DOMAIN.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 750 770 POTENTIAL.

Query Match	5.2%;	Score 311.5;	DB 1;	Length 2033;
Best Local Similarity	21.1%;	Pred. No. 1.7e-07;		
Matches 236;	Conservative 129;	Mismatches 331;	Indels 425;	Gaps 50;

RESULT 2

ID	HMCU DROME	STANDARD;	PRT;	2175 AA.
AC	P10180;	Q9W306;		
DT	01-MAR-1989	(Rel. 10,	Created)	
DT	01-MAR-1989	(Rel. 10,	Last sequence update)	
DT	28-FEB-2003	(Rel. 41,	Last annotation update)	
DE	Homeobox protein cut.			
GN	CT OR CG1387.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyigota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88232956; PubMed=2897632;			
RA	Blochlinger K., Bodmer R., Jack J., Jan L.Y., Jan Y.N.;			
RT	"Primary structure and expression of a product from cut, a locus			
RT	involved in specifying sensory organ identity in Drosophila."			
RL	Nature 333:629-635(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-Berkeley;			
RX	MEDLINE=20196006; PubMed=10731132.			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sturgeon G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Plannkoch C., Baldwin D.,			
RA	Abil J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,			
RA	Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.U., Wei M.-H., Ibbagan C.,			
RA	Jamali B.E., Kalush F., Karpen G.H., Ke Z., Kentonson J.A., Ketchum K.A.,			
RA	Kajim J.M., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasco P., lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McInerney D.,			
RA	Mervulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,			
RA	Palazzolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D., Scheibel F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,			
RA	Switskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			


```

Db      1093 LIARPRTAPPSPLFSPSLFGAAGMGASNA-----PAMADEMNRHVEFEI 1143
Qy      995 LPTFNRRNDYAA-----TSSAGSGEVKSGSGMGNYA 1028
Db      1144 AKLOOHQOOQAAQAQAFPFNSSLMALQOOVLNGAODLSLAAAKIKING----- 1196
Qy      1029 DIKRERLSADSGSGS---DEHSASHINNNSDLAHNNKSGGCG-----GGGGGNG 1076
Db      1197 --QSSLSHSGSSSCSGKDGERRDAPPS-----LHGKKSBBGGTPPAPPPSGGTAG 1249
Qy      1077 -----QTNG-NGKSSRMTSRDSETPDASFKSGENGQOONHKMMDLNGSSSSSHIKC 1128
Db      1250 APPTAPPTGAGSSNSAPSPLSNSILPPLASSQGEFFAATASPLQRM-----ASTNLI 1305
Qy      1129 ESEATGHSFGHHTTSLHEKLAQIKAEQVDQADQL 1165
Db      1306 TQPPVTPHSTPQRPRTKAV---LPITQOQFDMFNLI 1339

RESULT 3
BUN2_DROME STANDARD; PRT; 1211 AA.
ID      BUN2_DROME
AC      Q24523; Q9VK78; Q9VK79;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Bunched protein, class 2/class 3 isoforms (shortsighted protein).
GN      BUN OR SHS OR CG5461.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila..
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND
RP      FUNCTION.
RC      TISSUE=EYE-antennal disk; PubMed=7555710;
RX      MEDLINE=96038094; PubMed=7555710;
RT      "Shortsighted acts in the decapentaplegic pathway in Drosophila eye
RL      development and has homology to a mouse TGF-beta-responsive gene.";
RL      Development 121:2835-2845(1995).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).
RC      STRAIN=Berkley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abail J.F., Agbayani A., An H.-J., Andrews-Ffiankoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Bencs P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA      Botkova D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,
RA      Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jajala M., Katush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA      Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mactel B., McIntosh T.C., McLeod W.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,

```

```

RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
CC      -!- FUNCTION: Probable transcription factor required for peripheral
CC      nervous system morphogenesis, eye development and oogenesis. May
CC      be required for the transmission of the dpp signal and for a
CC      morphogenetic movement of the medulla in the brain that reorients
CC      the second optic lobe relative to the first. Plays a role in
CC      determining proper dorsal cell fates leading to the formation of
CC      the dorsal appendages.
CC      -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC      -!- ALTERNATIVE PRODUCTS:
CC      Name=Alternative splicing; Named isoforms=3;
CC      Comment=Experimental confirmation may be lacking for some
CC      isoforms;
CC      Name=Class 2;
CC      IsoId=Q24523-1; Sequence=Displayed;
CC      Name=Class 1;
CC      IsoId=Q24522-1; Sequence=External;
CC      Name=Class 3;
CC      IsoId=Q24523-2; Sequence=VSP_006670;
CC      -!- SIMILARITY: BELONGS TO THE TSC-22/DIP/BUN FAMILY.
CC      -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC      gene model prediction.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; LA2512; AAC41608.1; -; ALT_SEQ.
DR      EMBL; AE003636; AAF53200.1; -; ALT_SEQ.
DR      EMBL; AE003636; AAF53201.1; -; ALT_SEQ.
DR      PIR; T13804; T13804.
DR      HSSP; P80220; 1DIP.
DR      FLYBase; FBgn0010460; bun.
DR      InterPro; IPR000580; TSC-22_Dip_Bun.
DR      Pfam; PF01166; TSC22; 1.
DR      ProDom; PD007152; TSC-22_Dip_Bun; 1.
DR      PROSITE; PS01289; TSC22; 1.
KW      Transcription regulation; Nuclear protein; Alternative splicing.
FT      DOMAIN 15 31
FT      DOMAIN 76 86
FT      DOMAIN 97 102
FT      DOMAIN 237 241
FT      DOMAIN 249 254
FT      DOMAIN 261 265
FT      DOMAIN 306 321
FT      DOMAIN 322 328
FT      DOMAIN 348 356
FT      DOMAIN 607 619
FT      DOMAIN 661 668
FT      DOMAIN 743 746
FT      DOMAIN 759 765
FT      DOMAIN 801 817
FT      DOMAIN 817 822
FT      DOMAIN 832 838
FT      DOMAIN 884 891
FT      DOMAIN 927 947
FT      DOMAIN 1001 1005
FT      DOMAIN 1011 1014

```


RA Ballew R.M., Basu A., Baxendale U., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butts K.C., Busam D.A., Butler H., Caddieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gargliell A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.R., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Ralencat K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Silden-Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tecor C., Turner C., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
[5]
RN SEQUENCE OF 528-1073 FROM N.A.
RP STRAIN=Canton-S;
RC MEDLINE=96068664; PubMed=7479849;
RX Field G.J., Thummel C.S.;
RA "Isolation, regulation, and DNA-binding properties of three
RT *Drosophila* nuclear hormone receptor superfamily members,"
RL Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608(1995).
CC - FUNCTION: BINDS TO NGFI-B RESPONSE ELEMENTS. PLAYS AN IMPORTANT
CC ROLE IN LATE STAGES OF EPIDERMAL METAMORPHOSIS.
CC - SUBUNIT: FORMS A HETERODIMER WITH USP.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P49869-1, Sequence=Displayed;
CC Name=Short;
CC IsoId=P49869-2, Sequence=VSP_003714;
CC - TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN PREBLASTODERM
CC EMBRYOS, SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL
CC TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND
CC BRAIN COMPLEXES, BUT NOT IN OVARIES.
CC - DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.
CC HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.
CC SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN
CC LARVAE.
CC - SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).

CC EMBL, X89246; CAA61534.1; -;
DR EMBL, Y15606; CAA75690.1; -;
DR EMBL, AJ002073; CAA05172.1; -;
DR EMBL, AE003667; AAF53914.1; -;
DR EMBL, U36762; AAC46926.1; -;

```

Db 337 LSHEEDQLISNL-----TDSVSVSHSELSDFLF 365
Qy 473 -----AGIGAMPKDEQTQVDFIKGLLVANSIGIMEGLNL-----LSASQ 515
Db 366 PSDSNNLSLPTTSGVPPNPADLTSSINLTK--LTC-----LRDKRLSSIFEQLSSSQ 419
Qy 516 E-----NSNGNSALLL-----QQQHQHQQHQQHQQHQQHQA----- 549
Db 420 EQQLCLLSLRSSSDPAIALHAQOQQOQQOQQOQQOQQOQHQLQLDISPIGGLSCG 479
Qy 550 -----AYRHLPKSETPTETNSSLDPNDAEDPIIKTSPK-----VSGPASS 591
Db 480 SLPSPFOETYSLLKYVSSSGSSSPQOAS--SSTAAAPTPTQVTLTKMDEDCFPPLSGGMSA 537
Qy 592 SLSFGGLVGHGHHPLNNNSLSISNNSHNSNRHNSRNSRPHASPMILAAVAGCGV- 650
Db 538 SPPASQOL--QQLHTLQSOAQOMSHPNSSNNSNNGNSHNNS-----GGYN 581
Qy 651 -----SAGNSLLTSSSSSIQK-----MMASNIQRIQIESQGESLRNGVSDCSSNN 697
Db 582 YHGHNAINASANLSPSSASSLVEYNGVSAADNFYQ--QOQOQQOQVQGHNY---NSHN 637
Qy 658 GGSLSGLTKKPSIS-----VAKITGDTSPFGA-----SPNL 730
Db 638 GERYSLP-TFPTISLAAATAVAEAAATVGGPPVPRASLPVORTVSPAGSTAQSPKL 696
Qy 731 ----LSQOHSAN 739
Db 697 AKITLNQHSAN 709

```

RESULT 5

ASX_DROME STANDARD; PRT: 1669 AA.

```

ID ASX_DROME STANDARD; PRT: 1669 AA.
AC Q9V727, Q76930;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycomb protein Asx (Additional sex combs).
OS ASX OR CG8787;
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
CX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND DEVELOPMENTAL
RP STAGE.
RC TISSUE=imaginal disks;
RC MEDLINE=96146384; PubMed=9477319;
RA Sinclair D.A.R., Milne T.A., Hodgson J.W., Sellard J., Salinas C.A.,
RA Kyba M., Randazzo F., Brock H.W.;
RT "The Additional sex combs gene of Drosophila encodes a chromatin
RT protein that binds to shared and unique Polycomb group sites on
RT polytene chromosomes."
RL Development 125:1207-1216(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Bayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA Borkov D., Botchan M.R., Bouck U., Brockstein P., Broclic P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Galbraith W.M., Glasser K.,
RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasio P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPietron D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith F.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [3]
RP INTERACTION WITH TAN
RX MEDLINE=21290825; PubMed=11397012;
RA Dietrich B.H., Moore J., Kyba M., dosSantos G., McCloskey F.,
RA Milne T.A., Brock H.W., Krause H.M.;
RT "Tantalus, a novel ASX-interacting protein with tissue-specific
RT functions."
RL Dev. Biol. 234:441-453(2001).
CC -i- FUNCTION: Atypical Polycomb group protein, which may be involved
CC in both Polycomb group (PCG) and trithorax group (trxG) complexes.
CC PCG and trxG proteins act by forming multiprotein complexes, which
CC are respectively required to maintain the transcriptionally
CC repressive and transcriptionally active state of homeotic genes
CC throughout development. PCG and trxG protein complexes are not
CC required to initiate repression and activation, but to maintain it
CC during later stages of development. Both complexes probably act
CC via methylation of histones, rendering chromatin heritably changed
CC in its expressibility.
CC -i- SUBUNIT: Interacts with Tan.
CC -i- SUBCELLULAR LOCATION: Nuclear; associated with chromatin.
CC Colocalizes with many PCG sites on polytene chromosomes. It also
CC associates with many unique sites on polytene chromosomes.
CC -i- TISSUE SPECIFICITY: Highly expressed in nurse cells and deposited
CC in oocytes late in oogenesis. Ubiquitous in early embryos. Late
CC embryos show higher levels in CNS and neuroectoderm.
CC -i- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC Early embryos have high levels of expression, this drops off and
CC zygotic expression begins at 3-6 hour embryos. Expression levels
CC are low in larvae and medium in pupae and adults.
CC -i- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs, which may
CC be required for an association with nuclear receptors (by
CC similarity).
CC -i- SIMILARITY: Belongs to the Asx family.
CC -i- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -i- CAUTION: Ref.1 sequence differs from that shown due to
CC frameshifts in positions 608 and 719.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ001164; CA04568.1; ALT_FRAME.
CC EMBL: AE003814; AAF58239.1;
CC FlyBase; FBgn0000142; Asx.

```

DR PROSITE; PS01359; 2F PHD 1; FALSE NEG.
 DR PROSITE; PS50016; 2F PHD 2; FALSE NEG.
 KW Transcription regulation; Repressor; Nuclear protein; zinc;
 KW Metal-binding; Zinc-finger; Repeat; Developmental protein.
 FT ZN_FING 1632 1669
 FT DOMAIN 8 12
 FT DOMAIN 122 126
 FT DOMAIN 129 152
 FT DOMAIN 638 715
 FT DOMAIN 747 751
 FT DOMAIN 862 1202
 FT DOMAIN 1287 1290
 FT DOMAIN 1520 1524
 FT SITE 1527 1536
 FT SITE 224 228
 FT SITE 244 248
 FT CONFLICT 14 15
 FT CONFLICT 187 187
 FT CONFLICT 1253 1253
 FT CONFLICT 1520 1520
 SQ SEQUENCE 1669 AA; 179841 MW; F65D87398D67D321 CRC64;
 Query Match 4.5%; Score 272.5; DB 1; Length 1669;
 Best Local Similarity 19.4%; Pred. No. 9,4e-06;
 Matches 239; Conservative 160; Mismatches 449; Indels 387; Gaps 50;
 41 QPIAAGEDPSQY-----NHSSKEIQSNPNH-----CKTENHLE 78
 565 KPEALAPVDVANDQVSYLQVVELAETKAPLDNSNEADITGTNSHPVFSDDTIDHVF 624
 79 QOHNQSQLEEDSEENQTSNDSRTPTPGA-----TSTPSPPEPIDMPBSAKNFCV 132
 625 QEH-----QSTINNFSTSSSSNTATTANKLEHDKPEDSLPI----- 666
 133 NGRLLTVNAQGLVAESAATATSSSTNSHGHQSDNSNSASLPHHISSSSSNNSSG 192
 667 -----ASSISGSTPASTITSTS-----CTSSSSSSASMS-----SSCSSNSGSGT 707
 193 NRAHIAASARATPAATPANSLELYKLTQRAKMTSMGMAQLQAFSLADFNLIN 252
 708 TAPTTSSSAGAPTA-----LTIAAAETTLANVQ-----MLS 741
 253 SIASQOQOQOQOQIASAVPTTSEVSAAIISFALKDTSPSVDAPLDLSKPS-----PNS 308
 742 TVAKLQOQOQO-----ELPVELNSNEMVQVGHQD 769
 309 SISGDVKSVRACATPTPG--RRAVSEEDSRALODVYANKLDARKSASQHEORSIIDN 366
 770 WNFQGIK-----LSSSQSSGDQQRNLISHAID--LMDVY-----QDADVIDD 809
 367 RLFFKMKHHDQDHP--GDELEPSND-----DAEA-----EVDNSASTPV 404
 810 IMHNDVCHDVLCDEDEGQDEDEDEVEVCMTEBQQLIDSEBAVREIVDKLQOQOQOON 869
 405 YPAEFARQLKRLSHLSEHN-----GSDLGED-----VDRGSPKMGHPACGNASA 450
 870 QOQHQQHQLHIDVQGLAHSFMRQAHSEFGNDIGEMLCDAVRMAAEMEVSYITNMS 929
 451 NOGAPASIPLDANVLHTMLAAGIAMP-----KLDETQVGDPIKGLLVANSGIMN 504
 930 NSNDSN-----NISLCSSTNSLTINQPHQASQOPOQNAQOQOQIIVDSNGQIIG 984
 505 EGLNLLSAQENSGNSASLLIQO-----QOQHQQHQQHQQOQOQOQVAAVR---H 553
 985 NFLL-----QOQROQOQOQOQOQOFTLQAAAAQOQOQOQOQOQOQOQOQATSSNSLKG 1038
 554 RLPKSEIETNSLDPNDASEPILKIPSFVSGPSSSSLSIPGLVGHHPLNNNSL 613
 1039 TLPLALRNGTQOFLSPNLIAG-----QHQQOQOQOQ 1069
 614 SISNNSHNSHNRGNSRSPHSASPMIAAVALQGGYAGNSLLTSSSSSIQKM-----M 668
 1070 EGHQOQATAQCKHQIQOQFALQOQLHORQLLAQ---AANNVLQOQOQOQOQOQVALLPTTQ 1126

QY 669 ASNIQROINQSGQESLRNGVNS--DCSSNNQSSSLGKKPSISVAKIIGTDTSRFGAS 727
 DB 1127 AKFIAPKPNITIS---MTPRPNASPTTAATTANTASI-----PS-AVANVAVTGQOQOOSP 1178
 QY 728 P-----NLLSQOHNHSAHLTHQOQ-----OQQLSAQALGKGTTP 762
 DB 1179 FVPAPOQQTVOQOQLAHNSNMQOQLPVNLTMTKTLPSGVPFTTIAQOLRPMPTGKRKA 1238
 QY 763 KRGKRYVDRDSLVEAKAVQRGEMSVHRAGSYGVHSTLEYVKERHLMRPKRKEP-- 820
 DB 1239 TSNR-----LPPGAVNLERS---VOIQAVYIQ--NSPNEMLKAQLRPPAA 1279
 QY 821 ---KPODLVGLTGPANKLOIDKAKGPHGSKSLNALKNONQAAAAA----- 868
 DB 1280 ILNHQOFTTTAPRPIPVTLN---VSTVAATPMSNTTTATGSAALAAAPAPQNVLKOE 1336
 QY 869 -----AAAAAATENGKLPLEFAGPOLSFQPMF---WQTNATNAYGLDFNRITAM 921
 DB 1337 ELTVGAVGAGALPAGLPVNVMGV-----RPGVYKVIQPRMSG-----FPRKXYVQ 1383
 QY 922 RNPQASNHGMLKSAQDMVENVYDGIIRKTLQASEGNSAAGNSNGNGHGHG 981
 DB 1384 RKP-----SPTTLIRVF-----SPGEGATATQOQLQOHHQS----- 1419
 QY 982 HALLDQLVKKTPLPFTNHRNDYAATCSSASGESVKSQSPMGVYADIKERLSADSG 1041
 DB 1420 -----TSPVPVQNPQ-----QAPPELITONGN--QOYLVRANVGA-ADN 1459
 QY 1042 SSDEHSASHINNNSDLDHNKNSGSG--GGGGNGQOTNGNRSSRMTSRDSETDAS 1098
 DB 1460 QAPASSAPRNHQOQFVTVQNPPLHSINGIRPGGGRASVDYTAGSGNVLAIPISADAL 1519
 QY 1099 SFXSGENGQOQNMKMDLNGSSSSSHIKESSEA 1133
 DB 1520 HHHNHEMQOQOQOQOQOQOPLGVGVAANIVRNIAA 1554
 RESULT 6
 SUZ2_DROME STANDARD; PRT; 1365 AA.
 AC P25172;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Suppressor 2 of zeste protein (Protein posterior sex combs).
 GN SU(2)2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=91279476; PubMed=2057369;
 RA Brunk B.P., Adler P.N.;
 RT "The sequence of the Drosophila regulatory gene Suppressor two of zeste.";
 RL Nucleic Acids Res. 19:3149-3149 (1991).
 CC -!- FUNCTION: REGULATES EXPRESSION OF THE HOMEOBOX SELECTOR GENES BY INFLUENCING HIGHER-ORDER CHROMATIN STRUCTURE THROUGH INTERACTION WITH OTHER PROTEINS.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).


```

CC      -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMIUM AND
CC      ACCUMULATES IN NURSE CELLS DURING OOCYGENESIS. STORED IN THE EGG,
CC      BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC      PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC      -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: U11383; AAB60216.1; -
CC      EMBL: X59772; CAB36921.1; Alt_SEQ.
CC      PIR: A56038; A56038.
CC      HSSP: P07248; 2A0R.
CC      TRANSFAC: T00669; -.
CC      FLYBase; FBgn003028; ovo.
CC      InterPro: IPR007087; Znf_C2H2.
CC      Pfam: PF00096; ZF-C2H2; 3.
CC      SMART; SM00355; ZNF_C2H2; 4.
CC      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC      PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
CC      Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
CC      Transcription regulation.
CC      FT DOMAIN 62 66 POLY-ALA.
CC      FT 72 77 POLY-GLY.
CC      FT 80 85 POLY-GLY.
CC      FT 98 108 POLY-GLY.
CC      FT 144 152 POLY-HIS.
CC      FT 153 159 POLY-ASN.
CC      FT 336 339 POLY-GLN.
CC      FT 347 353 POLY-GLN.
CC      FT 357 361 POLY-GLN.
CC      FT 410 414 POLY-GLN.
CC      FT 426 432 POLY-GLN.
CC      FT 445 453 POLY-GLN.
CC      FT 456 459 POLY-GLN.
CC      FT 466 474 POLY-GLN.
CC      FT 497 517 POLY-ALA.
CC      FT 524 529 POLY-SER.
CC      FT 549 558 POLY-ALA.
CC      FT 639 651 POLY-ALA.
CC      FT 717 725 POLY-ALA.
CC      FT 797 802 POLY-GLN.
CC      FT 820 823 POLY-GLN.
CC      FT 826 832 POLY-GLN.
CC      FT 874 896 C2H2-TYPE 1.
CC      FT ZN_FING 902 924 C2H2-TYPE 2.
CC      FT ZN_FING 930 953 C2H2-TYPE 3.
CC      FT ZN_FING 969 992 C2H2-TYPE 4.
CC      FT CONFLICT 647 647 A -> R (IN REF. 2).
CC      SEQUENCE 1028 AA; 110620 MW; D7068BBB2BC0FF677 CRC64;

Query Match 4.3%; Score 259; DB 1; Length 1028;
Best Local Similarity 20.4%; Pred. No. 2.3e-05;
Matches 202; Conservative 122; Mismatches 362; Indels 304; Gaps 45;

```

```

QY 330 AYSEEDLSRALQDVYANKLDRKASQHEQRSILNRLFMKAKHDOEDHDEDSN 389
DB 255 DHELEQSDAND--LSSTL-----QRCVD-----DEESTPRQDFE--- 289
QY 390 DDAEAEVDSNASTP--YPAEFARAQLRKLSHLESHNGSDIGEDVDR-GSPKRGHRPACG 446
DB 290 -----LVSTPRLTPDSTVPYEQHNTNTQTLDVLHENLTLQTLNHLVIGSGSQOQHQQH 344
QY 447 NASANQCARAPRIPDANVLHTLMLAAGIGM-----PKDEYQYGDFTKGLLVANS- 499
DB 345 GVQQOQOQOQHVVQOQOQ--HNVOQOQGVQOQHVQOQPPPSYQAT-----RGLMQOQP 396
QY 500 --GIMNEGGLNLTSAQENSGNASLLQOQH--QQHQQHQQOQOQOQHVAAV-RH 553
DB 397 QHGGVQOQ-----AATMSQOQOQLLQOQOQSHQOQOQOQOQHAAVQOQ 440
QY 554 RLPKSETPETNSSLDPDASEDPLIKTPSEKVSGRASSSSLSPLGVGHHPLNNNSL 613
DB 441 NIYAQOQOQOQOQHQQOQH-----HFHHQOQOQOQPQ 472
QY 614 SISNNSNHSNHSRNGSNRSPHSASPMLLAAVAGVYAGNSLLTSSSSSIQKMMASNIQ 673
DB 473 POSHSHHHGHGHDNSNMSLP--SPTAAALAAAAALAAAAALAAAA--AAHLQ 520
QY 674 RQINEQSGQESLRNGVNSDCSSNNGSGSSSLGKKPSISVAKIIGTPTSRFGASPNLSQ 733
DB 521 RPMSSSSSGGTNSNSGSSN--SPLDANAAALAAAL--DT-----KPLQSV 569
QY 734 QHSAHHLTHQOQOQOQLSAGELGKTRPKRGKRYNDRDSLVAVAVAGVEMSVHAG 793
DB 570 SNPIGQPLNTQSOQOQOQOQITLMKTR--YTER--VEMV----- 605
QY 794 SYGVPSHSTLEYKVKERHLMRPRKRPKOPDVLGLGPNANKLQLDKLYKAGPHGSKLSN 853
DB 606 -----SMDVTVKPELF-----SELKPEMTETI--AELTLEA----- 635
QY 854 ALKNQNNQAAAAALAAAAAATPNGIKLPLFEAGFOALSFCQPMFQPTNATN----- 908
DB 636 -----ETTAALAAAAALAAAAAATTSATE-----GTVLAANA--PLSSGKLGRA 680
QY 909 ---AYGLDFNRITAMR-NPQA---SNHGLMKSADWENVYDGIIRKTLQASENGS 960
DB 681 KAVAYGSTMITLISLTKSSPEVPATKTVHRTTSLATAAATAGGLASPTVSVLNES 740
QY 961 AAGGNSG-----SNGNGH-----HGCHGHALDLDLVLKKTPLPTNHRNN 1003
DB 741 KVLORLGLRPDLQLEFVNG-GHGIKNPLAVENAHGHHRI-----RNI 783
QY 1004 DYAAATCSASGES--VRSQSPMGNVADIKERLISADSGS-SDEHSASHINNNSDLA 1060
DB 784 DCIDDLKHHGHSQHQOQOQSPQOQNMQGVQOQSVQOQSLQOQOQOQHNNHNSAS 843
QY 1061 HNKXKSG-----GGGGGGGNGOTNGNR 1083
DB 844 SNASHSGSAEALCMWSSGANEDESSGNXK 873

RESULT 8
HLES_DROME STANDARD; PRT; 1077 AA.
AC Q02308; Q9VDK0; Q9VDK1;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hairless protein.
GN H OR CG5460.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).

```


RA Bang A.G., Posakony J.W.; PubMed=15116831;
 RT "The Drosophila gene Hairless encodes a novel basic protein that
 RL controls alternative cell fates in adult sensory organ development.";
 RN Genes Dev. 6:11752-11769(1992).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=93041287; PubMed=1419950;
 RT Maier D., Stumm G., Kuhn K., Preiss A.;
 RL "Hairless, a Drosophila gene involved in neural development, encodes
 Mech. Dev. 38:143-156(1992).
 [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blake R.G., Champagne M., Pfeiffer B.D.,
 RA Abul J.F., Agbayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwatz R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Botkova D., Botchan M.R., Bouck H., Brokstein P., Brothier P.,
 RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodaira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattali Y., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svaykdas R., Teecor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [4]
 RP FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
 CC DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
 CC FATES BY THE TRICHOGEN (SHAFT) / TOMOGEN (SOCKET) SISTER CELL
 CC PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
 CC OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
 CC TOMOGEN FATE.
 CC
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=002308-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=002308-2; Sequence=VSE_006952;
 CC Note=No experimental confirmation available;
 CC
 CC -1- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
 CC DISCS.
 CC
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

[illegible]

Qy	467	TTLMILAAIGAMPKLDIEQTQGDPIKGLLVANSGGIMNEGILNLTLSAQENSGNAGSL	526
Db	525	KTSAVIA-----STTSSDR-----GGLSHALTH-KVSPESATDAGLV	565
Qy	527	QOOQHOOHQHOOOQOOOHVAAVRRLL-----PKSEPTENSLDND	571
Qy	566	--EYHTQVSPRRKILREFEKVSLIEDNCVNNSGAGSSGAGKRSRAKGTSTSPAGK	623
Qy	572	ASEDPILKIPBFKUSGP--ASSSLSGGLVGHHHPLNNNLST-----SNNSHSS	623
Db	624	AS--PMNLAPQOKRPSPPGSSSSSTSPATL--STGTPRLNSSYSIHSLLGGSSGSS	678
Qy	624	NSHRNGSNRSPHASPTLAAVAOQGYDAGNSLITSSSSSIQJMAWNIQROINOSGOE	683
Db	679	SSSSSGKKCGDH-----PALIISVHHIQRHSMYQSPSSSYPRAL-----	717
Qy	684	SLRGNVNSDCCSNNNGS---SSLGKYK-PSISVAKTI-----GTDTSRFG	725
Db	718	-LTSPKSDVAGSNGSGGKSPSHGTCKRSPYSAGSFVDYGHSEFYPDPYAGARPTSG	776
Qy	726	ASPMLISQOHNSAHLTHQOOQOOLSAQELGKGRPRKGRKRYNDRDLSLEAVAYORG	785
Db	777	SASQDLSPPRS-----PAPRATPTRTVPKK-----TASITRE	809
Qy	786	EMSVHACGSYYGVPHSTLEYKVKERHLMRPRKREKPOPDVLGILGPANKLQDLXKAP	845
Db	810	FASPSASSSSGSPS-----GDRSASPRERHHQOOPHLQRSSPLHYWMPPEPVNG	861
Qy	846	HGGSKLTSALKNONNQAAAAAIAAAAAAATP-----NGCLKLPLEA	888
Db	862	NGSASPSPTASPTSSHHAAVAAAAAIAAIAIPSPSTNPYITSLAALRHPLMHNHQT	921
Qy	889	GPQALSFQPNMFQPTNATNAYGLDFNRITEAMRNPQASNNHGLKMSAQDWENVYDGI	948
Db	922	GASPL-LSPH---PQPGSAAAAA---AAAAARLSPOSAVH-----AFAYNGVG	963
Qy	949	RKTIQASGNGSAAAGNGSNGNCH---GHCHGHALLDQLLVKTPLPFTNRRNDY	1009
Db	964	AAVAAAA---AAAAFGQPADBPYHHPHLAHHQPH-----PAALTTHSPAH	1008
Qy	1006	AAT-----CSSASGESVTKSGSPMGVADIKRRLSLADSGSSDEE-HSASHINNNS	1055
Db	1009	LATPLTDSSTDQMSATSHRRASTSPSSSASASASASATGSASSAMFHTSLNDESS	1066
Qy	1058	DLAHKMK	1065
Db	1069	DLPLNLK	1076

ID	PROS_PROVI	STANDARD;	PRT;	1556 AA.
AC	Q9UEA1;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Protein prospero.			
CN	PROS.			
OS	Drosophila virilis (Fruit fly).			
OC	Neoptera; Melezaa; Arthropoda; Insecta; Pterygota;			
OC	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephyndroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7244;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RK	MEDLINE=20503846; PubMed=11051550;			
RA	Xu C., Kaufmann R.C., Zhang Y., Klady S., Carthew R.W.;			
RT	"Overlapping activators and repressors delimit transcriptional			
RT	response to receptor tyrosine kinase signals in the Drosophila eye.,"			
RL	Cell 103:87-97 (2000).			
CC	-I- FUNCTION: Required for proper neuronal differentiation of most or			
CC	all neurons and their precursors in central and peripheral nervous			
CC	systems, axonal outgrowth and pacifying. Not required for the			

CC	specification of neuronal identity. May regulate transcription by
CC	binding to DNA (By similarity).
CC	-1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC	-1- SIMILARITY: BELONGS TO THE PROSERO HOMEOBOX FAMILY.
CC	-1- SIMILARITY: Contains 1 homeobox domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC	or send an email to license@ebi.ac.uk).
CC	-----
DR	EMBL; AF190405; AAF06660.1; -
DR	FLYBase; FBgn028753; Dv1r.pros.
DR	Pfam; PF05044; Proxl; 1.
KW	Nuclear protein; Transcription regulation; DNA-binding;
KW	Homeobox; Developmental protein.
FT	DOMAIN 4 12
FT	DOMAIN 19 36
FT	DOMAIN 150 172
FT	DOMAIN 206 209
FT	DOMAIN 237 292
FT	DOMAIN 305 309
FT	DOMAIN 349 381
FT	DOMAIN 443 473
FT	DOMAIN 544 547
FT	DOMAIN 574 1080
FT	DOMAIN 888 906
FT	DOMAIN 1027 1030
FT	DOMAIN 1045 1054
FT	DOMAIN 1057 1062
FT	DOMAIN 1132 1189
FT	DOMAIN 1140 1145
FT	DOMAIN 1154 1163
FT	DOMAIN 1183 1189
FT	DOMAIN 1090 1097
FT	DOMAIN 1330 1337
FT	DNA BIND 1394 1456
FT	DOMAIN 1457 1556
FT	SEQUENCE 1556 AA; 171029 MW; 6FEACFEA2D73B644 CRR64;
SO	SEQUENCE
Query Match	4.1%; Score 249.5; DB 1; Length 1556;
Best Local Similarity	18.7%; Pred. No. 0.00011;
Matches 232; Conservative	167; Mismatches 434; Indels 405; Gaps 52;
QY	47 GSEDEPGYNSKSEIKSQSNPNHCKTENHRLQOHNGSQLLEEDSENQTSHSSRTPT 106
DB	348 GSSNSSSSSHSNSSNNISAS-----NSGLRRKSSDSDMSQREBRA--T 389
QY	107 PGATSTSPPEPPIIDWRPSAKCNCFVAGRLTLTVAQG--KLVAESAATATSS----- 157
DB	390 PATTVEPOLQNEVLPYTKKEAVDDMLDEVELLGLSGHGRNSDLESLASPSHSDMLLDN 449
QY	158 -----TSNSHTHQHSDSDNSASLPHHIISSSSSSNNSSG-----NRAR--HIAAA 201
DB	450 SKDEVLDNDIILEDDEDDDDDD--DCIETRGESNTSGSNLTKKPKMELKRAVEHI-VS 506
QY	202 SARATPAAPAFNSLELYKLITQRAAYKTSMDMAQLAQPSLLADFNILNSLSQQQQQ 261
DB	507 SMRCSPTATQAQGLQV-----NGCKKRKP-----YQPOQH 537
QY	262 QQQQIAAVPTTSEVGAASAIAPLAKDTPSPVDAPIDLSSKSPNNSISGDVKSVRACA 321
DB	538 AMERYVAAAAGLNFGLNLQSMWLDQEDSESNELESP-QIQQKVEKALKLSQLRSM--- 592
QY	322 TPTPSGRRAVEEDL-----SRALDDVANKLIDARKASQHHIEPSSIIDNRFLFX 370
DB	593 -----QEQLAMQCKYVGLGSRMEDESCQELD-----NEHQDQMDMSMEQ--Q 634
QY	371 MKHHDDQDHGDGDELESDNDAAEAVDSNASTPVYPAEFAFARQLRKLSH--LSEHNGSD 427


```

CC Name=5; Synonyms=BCORE-NS-23;
CC IsoId=001295-5; Sequence=External;
CC - DEVELOPMENTAL STAGE: ACCUMULATES TO A HIGH LEVEL AT THE BEGINNING
CC OF THE ECYCLONE RESPONSE, DURING THE METAMORPHOSIS OF IMAGINAL
CC DISKS IN PUPF STAGE 1, AND ABRUPTLY DISAPPEARS AFTER SEVERAL
CC HOURS.
CC - INDUCTION: INDUCED AS A PRIMARY RESPONSE TO 20-HYDROXYECYCLONE IN
CC THIRD INSTAR LARVAL IMAGINAL DISKS.
CC - SIMILARITY: Contains 1 BTB/POZ domain.
CC - SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC - CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC TO 619 AND 656 TO 694 DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U51585; AAB09760.1; ALT. FRAME.
CC DR EMBL, AL009146; CAA15627.1; -.
CC DR EMBL, AL003421; AAF45647.1; -.
CC DR TRANSFAC; T01480; -.
CC DR Flybase; FBgn0000210; bf.
CC DR InterPro; IPR000210; BTB_POZ.
CC DR InterPro; IPR007087; Znf_C2H2.
CC DR Pfam; PF00651; BTB; 1.
CC DR Pfam; PF00096; zF_C2H2; 2.
CC DR SMART; SM00355; ZNF_C2H2; 2.
CC DR SMART; SM00355; ZNF_C2H2; 2.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
CC DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.
CC DR PROSITE; PS50097; BTB; 1.
CC KW Nuclear protein; DNA-binding; Developmental protein;
CC Zinc-finger; Metal-binding; Alternative splicing.
CC FT DOMAIN 32 97
CC FT ZN_FING 710 733 C2H2-TYPE 1.
CC FT ZN_FING 740 763 C2H2-TYPE 2.
CC FT DOMAIN 203 207 POLY-ALA.
CC FT DOMAIN 265 268 POLY-ASN.
CC FT DOMAIN 458 466 POLY-ASN.
CC FT DOMAIN 584 589 POLY-PRO.
CC FT DOMAIN 618 621 POLY-ALA.
CC FT DOMAIN 698 803 POLY-ALA.
CC FT DOMAIN 821 833 POLY-ALA.
CC FT DOMAIN 862 867 POLY-GLN.
CC FT CONFLICT 436 436 G -> D (IN REF. 1).
CC FT CONFLICT 621 621 MISSING (IN REF. 1).
CC FT CONFLICT 624 624 AV -> R (IN REF. 1).
CC FT CONFLICT 661 662 AV -> L (IN REF. 1).
CC FT CONFLICT 678 678 MISSING (IN REF. 1).
CC FT CONFLICT 722 723 KL -> NV (IN REF. 1).
CC SQ SEQUENCE 880 AA; 92305 MW; 500C04A3B665AAAF CRC64;

Query Match 4.1%; Score 246.5; DB 1; Length 880;
Best Local Similarity 20.0%; Pred.No. 7.6e-05;
Matches 202; Conservative 114; Mismatches 353; Indels 343; Gaps 41;

37 IEBOQPIADSEDPQYNNHSSKSIQSNPNH-----CKTENHRLDEQHNQSOLLLEED 91
Db LKTAIEVLNVSGIGTQOQADDTSHLAQIIONLANSRGRTPLNTHQSLRPHPHGS--LHDDG 161

92 SENNOTSHDSRTPPGATSTPSPRPPEPIDMRPSAKKFCVNGRLTVNAQSKVAESGA 151
Db GSSTLFSHQGAGSPPT-TAVPSPDSH-----INQLLKMA--MHRSSA 203

152 TATSSSTNSNHHQHDSDNSNSASLPHHIISSSSSSNNNSGNGRAHIIAASARATPAAT 211
Db AAAAEETSHAPRLRGSDN---SLPLSGAVGSSGNNNSPDLPLHARSASAGQTF----- 255

212 PANSLELYKLTORAAKRTSMDMAQAQSLADPNLINSLASQOQQOQQOQIATAVT 271

```

Dd		256	-----		ADPFTTIGHNNNN--	268		
Oy		272	PPTSEVSAAAISPALKDTPSPVDAPLDLSKRPSSISG-----		DYKSVACAPT	324		
Dd		269	-----	: : :	NPTGTNGSNGNGNCSANGNCISIDTKSGILTPS	314		
Oy		325	PSGRRAVSIEDLSRALQDVVANKLDAKYSAGHNEORSILDNRLLPKMHNDQEOHDGD-			383		
Dd		315	PLARA--GADVVKSEPMDCSNNA--NANDEHSND-----		-TGEHANRSSGGDG	362		
Oy		384	---ELDSNDNAAE--VDSNASTPYU---PAEFAAQLRKSHISEHNSGLSDGVDRGS			436		
Dd		363	GKSGLSSGNDEIGDDLASHHAAPRINSPE--NKMFHAAFPRNI--			408		
Oy		437	PKMGRHPACGMSANGAPASIPLDVANVLTHTMLAAG-IGAMPKLDETQTVCDFIKGL			493		
Dd		409	-----DPALLG-----LNTQLQOSDLAVSRQGST--GSLSGVI			443		
Oy		496	V-ANSQGINNEGVLNLISAQENSGNASLLIQOOHQHHQQHQQOQQOQHVAAVHR			554		
Dd		444	VPGSGCGTPS-----NSSNNNNNNNSV--NQOKVEQOSSPHQLQQQH--			486		
Oy		555	LKPSEPTEMTSLDPRDAEDPI LKI PRKYVSGPSSSSLSRGLVGCHNRLPNNNSSL			614		
Dd		487	----STPHNTS-----POLKEOPKS-----GG-----			505		
Oy		615	ISNNSSHSSNSH-RINGSNRS PHSAS PMLAAVAOCYGAGNSLTTSSSSIOKMAGNIQ			673		
Dd		506	--GCCKSSDLHI AASSERSLSRS--QMPRAGSHATPRTYAIVNGERE			554		
Oy		674	RQINEQSGQESLRNGVSDCCSNNGCSSSLGKKPISIVAKI IGTDTSRFGASP-NLLS			732		
Dd		555	REPERERERERERSLDERDLERPCTGSPRRPPS-----NHSHFGQRLSLP			604		
Oy		733	QQR--HSANH-----LTHQOOOOLSAQALGKTRKRGKYKNYRDLSLVELVAYQR			784		
Dd		605	SHQLATHNELLSAAAHHAHAAHAAHAAHALARAGSMENNHLNHRASLSPSGAVSS			664		
Oy		785	GEMSVHAGSYGV--PHSTEYUKYKER-----HLMFPRKREPORDVLGLGPANKL			836		
Dd		665	AASGAGRGGAGCGPGSGSLSVRADVDQAORLLRLPLMACHRCDVCGKL-LSTKL			722		
Oy		837	QLDKLYAGRPH-----			846		
Dd		724	TLKRHKEQQLDRLNNAVCLCHKVFRTLNSLNHKSIYHRQKXNHSYFNHGAVSQAG			783		
Oy		847	-GGSKLSNALKONNOAAAA-----AAAAAAAAAATPRGKLPFEA			888		
Dd		784	SPOSRIHQSISSLSAAAAAANNNSVVNGGSGVGAAGNVAAAAAAAMELLSLPTVGA			843		
Oy		889	GPOALSFPQMFWRQTNATNAVGLDFNRTTEAMRPOASNHGKLSKMODV			940		
Dd		844	AAVA-----GGTASSTL--QLAAHQOQOQSSPGIVKRCMDFL			880		
<hr/>								
RESULT 12								
E74B_DROME								
ID	E74B_DROME	STANDARD;	prt;	883 AA.				
Ac	P11536; Q9VVI7;							
Dt	01-OCT-1989 (Rel. 12, Created)							
Dt	01-OCT-1989 (Rel. 12, Last sequence update)							
Dt	15-SEP-2003 (Rel. 42, Last annotation update)							
De	Ecdysone-induced protein 74F isoform B (ETS-related protein E74B).							
Cn	E174AF OR E74 OR CG6285.							
Os	Drosophila melanogaster (fruit fly).							
Oc	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pelecygota;							
Oc	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;							
Oc	Ephydroidea; Drosophilidae; Drosophila.							
Ox	NCBI_TaxId=7227;							
Nn	[1]							
Rp	SEQUENCE FROM N.A.							
Rx	MEDLINE=89315191; PubMed=2501755;							

RA Janknecht R., Taube W., Lueddecke H.-J., Pongs O.;
RT "Characterization of a putative transcription factor gene expressed
in the 20-OH-ecdysone inducible puf1 74EF in Drosophila
melanogaster.";
RL Nucleic Acids Res. 17:4455-4464(1989).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=90199900; PubMed=2107982;
RA Butris K.C., Thummel C.S., Jones C.W., Karim F.D., Hogness D.S.;
RT "The Drosophila 74EF early puf1 contains E74, a complex ecdysone-
inducible gene that encodes two ets-related proteins";
RL Cell 61:85-99(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chamee P., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beron P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Dommes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell T.J., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Steplston M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [4]
RP DEVELOPMENTAL STAGE.
RX MEDLINE=94038699; PubMed=8223281;
RA Huef F., Ruiz C., Richards G.;
RT "Puffs and PCR: the in vivo dynamics of early gene expression during
ecdysone responses in Drosophila".;
RL Development 118:613-627(1993).
RN [5]
RP SUBCELLULAR LOCATION: Nuclear.
RX MEDLINE=118-613-627(1993).
RN [6]
RP ALTERNATIVE PRODUCTS:
RX Event=Alternative products; Named isoforms=2;.
RN [7]
RP Name=B; Synonyms=E74B;
RX IsoId=PI1536-1; Sequence=displayed;
RN [8]
RP Name=A; Synonyms=E74A;
RX IsoId=E20105-1; Sequence=External;.
RN [9]
RP DEVELOPMENTAL STAGE: In mid instar larvae salivary glands levels
increase during 86-94 hours of development and represent the
predominant isoform during puf1 stage 1. Levels remain relatively
constant in late larvae until the premetamorphic pulse of
ecdysone. Transcripts are detected after from puf1 stages 12-14

```

CC and 17-21.
CC -I- INDUCTION: THE EXPRESSION OF THIS PROTEIN IS DEVELOPMENTALLY
CC REGULATED AND IS CORRELATED WITH THE 20-OH-ECDYSONE INDUCED
CC ACTIVITY OF PUPF 74EF.
CC -I- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15087; CAA33195.1; -.
CC DR EMBL; M37083; AAA28494.1; -.
CC DR EMBL; AE003523; AAF49324.1; -.
CC DR PIR; S04722; S04722.
CC DR HSSP; P14921; 25MT.
CC DR TRANSFAC; T00210; -.
CC DR FlyBase; FBgn0000567; Eip74EF.
CC DR GO; GO:0006914; P:autophagy; IMP.
CC DR InterPro; IPR000418; Ets.
CC DR InterPro; IPR002341; HSF_ETS.
CC DR Pfam; PF00178; Ets; 1.
CC DR PRINTS; PR00454; ETSDOMAIN.
CC DR SMART; SM00413; ETS; 1.
CC DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
CC DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
CC DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
CC KW Nuclear protein; Transcription regulation; DNA-binding;
CC Alternative splicing; Developmental protein; Polymorphism.
CC FT DOMAIN 281 759 ALA/GLN/SER-RICH (ACIDIC).
CC FT DOMAIN 760 883 ARG/LYS-RICH (BASIC).
CC FT DNA_BIND 787 869 ETS-DOMAIN.
CC FT DOMAIN 27 51 POLY-SER.
CC FT DOMAIN 72 79 POLY-ALA.
CC FT DOMAIN 289 299 POLY-ALA.
CC FT DOMAIN 372 378 POLY-GLN.
CC FT DOMAIN 406 415 POLY-GLN.
CC FT DOMAIN 434 437 POLY-GLN.
CC FT DOMAIN 486 489 POLY-SER.
CC FT DOMAIN 510 524 POLY-ALA.
CC FT DOMAIN 614 630 POLY-GLN.
CC FT DOMAIN 639 642 POLY-GLN.
CC FT DOMAIN 651 661 POLY-GLN.
CC FT DOMAIN 695 703 POLY-ALA.
CC FT DOMAIN 707 711 POLY-SER.
CC FT DOMAIN 712 717 POLY-ALA.
CC FT DOMAIN 721 733 POLY-ALA.
CC FT VARIANT 867 867 H -> Q.
CC SQ SEQUENCE 883 AA; 94820 MW; 148D5031A18D1409 CRC64;

Query Match 4.1%; Score 244.5; DB 1; Length 883;
Best Local Similarity 20.4%; Pred. No. 9.6e-05;
Matches 177; Conservative 117; Mismatches 364; Indels 209; Gaps 28;

145 LVAESATP-ATSSSTNSHIOHSDNSASALPHHIISSSSSSNNNSGGR--ARHTAA 200
Db 7 LVAASAINFASQAALVNVSSSSSSSSSSSSSSSSSSSSSSLSATTPPAASVPTP 66
QY 201 AS--ARATPAATPANSLELYKLLTORAAKMTSDMAAQLAQSLLADENLIINSLSAQ 257
Db 67 TSPPPAAAAPRAEASPPAGAEIQE-----DGOQAKTQDPTMKQDMLEKTRQE 114
QY 258 QQQQQQQQQLASAVTTTSEVSAALISPAKLDTPSPSVADAPLDLSSKSP-----NSSI 310
Db 115 VKDEVNVEEPQAIYDTESVVARQSPSVASTKYPESELE--EISNKSPPVQDEDEESSESV 171
QY 311 SGDYKSVRACTPTPPSGRRAY--SEEDLSRLQDVVANKLIDARSAQSHHQRGILD--N 366
Db 172 ASDREKRVLYNHLRQQQHHSPSPPKTRSLTDVSKILWERK--QQLORSSVITNAP 228

```

```

QY 367 RLFFKMHDDQEDHDELESDNDPAEAEVDNASTPVVPAEFARQJRLKLSLSEHNGS 426
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 TLQPOHQOQPMSDIDEDETLFEDVDADADAVEADADEBELLEY-----QNGY 275
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 DLGEVDKSGSPMGHNPACGNASANOAGAPASIPLDANVLLHTLMLAAGICAMPKIDETOT 486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 DSPPLSLSLG-----GATSAASAASAAA-----ASAASRRGRGTYSGT 311
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 487 VGD-----FIKGLLVANSGGIMNGLNLISASQENSGNSALLQOQOQOHHQ 536
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 ESDSDACERAMRLPERKARSAAYKKLMKRYTTEIPYKOSTSPAPQOQLQOQHLL 371
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 537 QHHQOQOQOQH-----VAAVRHRLPKSETPET-----NSSLD 568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 Q--QOQOQOQPHNGSTFAGATALLHTKTEGNTLLTTLQLOQOQOQOGLGAANGSSNG 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 569 PNDASEDLPLKIPSRVSGPSSSLSPGGLVGGHHPLNNNSLSISNNSHSSNHRN 628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 NNAHQOQOQPLAIP-----QRPLHLNLSLGGALNPHHRNTTATTGSPSPADSGVSDVD 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 629 GSNRS-----PHSASPMIAAAGGYSGNSLLTSSSSSIQKMAASNIQ 673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 SSSSGGQPCADELKLKRLGMPATTSASAAAAAAAHHTGFLEHNLVQNNAAASL 545
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 674 RQINEQS-----GOESLRNGNYSDCSSNGSSSLGKKPKSISVAKIIGTDTSPRG 725
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 546 RNIMRSGVDPNRYGSSGAGSGG--PGRGNGNQTREYTLTTSFNAPATAAASQRG 603
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 726 ASPNLLSQOHSAAHLTHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 785
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 604 TTIN-----GYHSLHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 786 EMSVSRAGSYGVPHSTLEYKKEKHLMPRKREKRPQDVLGLTP-ANKQLDKLAKG 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 629 QQLAAHQOUSHQ--QQOALHQOUSHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 845 PHGG-----SKLSNALKNQNNQAAAAAAATPNCGLKPLFPAAGPOA 892
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 PHAGCHTSTIAAAAAAASVSSSSSAVAAAAAASAAAAAATAA-----AAGSGO 737
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 893 LSPQNMFMPTQNTATNAYGLDFNRITE 919
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 738 SVIQP-----ATSSVSYDLSYMLE 756
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13

```

SIMA_DROME STANDARD; PRT: 1507 AA.
AC Q24167: Q9VAA5;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Similar protein.
GN SIMA OR CG7951.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96269413; PubMed=8682312;
RA Nambu J.R., Chen W., Hu S., Crews S.T.;
RT "The Drosophila melanogaster similar bHLH-PAS gene encodes a protein
RT related to human hypoxia-inducible factor 1 alpha and Drosophila
RT single-minded.";
RL Gene 172:249-254 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RC MEDLINE=2019606; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang O.X., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blasej R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brodtier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiter K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.F., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
CC -1- FUNCTION: POSSIBLE DNA-BINDING TRANSCRIPTIONAL ACTIVATOR.
CC -1- SUBUNIT: Efficient DNA binding requires dimerization with another
CC BHLH protein.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN THE EMBRYO.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43090: AAC47303.1; -
DR EMBL: AE003772; AA57008.2; -
DR PIR: JC4851; JC4851.
DR FlyBase: FBgn0015542; sima.
DR InterPro: IPR001092; HLH_basic.
DR InterPro: IPR001610; PAC_
DR InterPro: IPR000014; PAS_domain.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HLH_1; 1.
DR PROSITE: PS50112; PAS; 2.
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW Activator; Coiled coil.
FT DNA_BIND 72 85
FT DOMAIN 86 126 BASIC DOMAIN.
HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

```

Query Match	4.0%	Score 243	DB 1	length 1507
Best Local Similarity	18.6%	Pred. No. 0.00021		
Best Match 241	Conservative	157	Mismatches 433	Indels 462
				Gaps 48

Qy	1	ERVAECCGRQWQKHQDKLTCSHINIEQO-----	-PIAIGSEDEBPQY	55
Db	461	ELIAQETKYEVTNPPIHSELQAKPQLQSEKKEKTEETKTATIPVTAISTAQO----		516
Qy	56	NHSSKEISQSNPNH-----	CKTENRLEQOHNSQLE-----	EEDSENNQTSH 99
Db	517	---IKQLESPNYPKQILQAEILLIKKENHSPGRITTAQLDSSSSSGLAPEERPPSYVA-		572
Qy	100	DSRTPTFGATSTPSP		115
Db	573	-SVLRSPAPRLPTPEPTAVLCKKPTLGYEPNLPPTTATAIISSSNQOLQIAQQTOLQ		631
Qy	116	PEPIDWRPASKCN-FCVNGRLT-----		YNAQ 143
Db	632	FQQRQDMSSKFCSLFADDGKGLTYLKEEPDDLHHLASTNCTQIDENTPPSDMVLGML		691
Qy	144	K-LVAESAATATSSSTNSHIOH-DSDSNSASLPHHISSSSSNNNSGGRARHIAAA		201
Db	692	TCLRPEDINSLDSTTCSFTASQHYQSPSSSSTAP---	SNTSSNNSYANSP--LSPL	745
Qy	202	SARAPRAATPANSLELYKLLTORAAKTSMSMAQLAQFSLDLFNLINSLASQOQOQ		261
Db	746	TPNSTATASNPSHQO-----		QOHNQOQOQ 771
Qy	262	QOQOQ-----	IASAVPT--TSEVSAALSPALKDTSPSPVAPLD--LSKRS	305
Db	772	QOQOQHHPQHHDNNSSSNIDPLFYVRESNNTSCQH-HSPSTSKSPBESSLPJLCS		830
Qy	306	PNS-----	SISGDVSVRACATPTPSGRRAVSEDL--	SFALQDVVANKLDAKASQ 356
Db	831	PNLSLQEBDDFSGTEAETAMRAPYRIPIDDMPLLETILIMCSPREDLQTNVPEKIDALQOQLO		890

[illegible]

RA MEDLINE=92069760; PubMed=1720353;
 RA Vaesele H., Geell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
 RA "Prospero is expressed in neuronal precursors and encodes a nuclear
 RT protein that is involved in the control of axonal outgrowth in
 RT Drosophila.";
 RA Cell 67:941-953(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92171948; PubMed=1540176;
 RA Matsumaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;
 RT "Cloning of the Drosophila prospero gene and its expression in
 RT ganglion mother cells.";
 RA Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93083413; PubMed=1842358;
 RA Chu-Lagrange O., Wright D.M., McNeill L.K., Doe C.O.;
 RT "The prospero gene encodes a divergent homeodomain protein that
 RT controls neuronal identity in Drosophila.";
 RA Development Suppl. 2:79-85(1991).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM C).
 RA MEDLINE=20503846; PubMed=11051550;
 RA Xu C., Kaufmann R.C., Zhang J., Klady S., Carthew R.W.;
 RT "Overlapping activators and repressors delimit transcriptional
 RT response to receptor tyrosine kinase signals in the Drosophila eye.";
 RA Cell 103:87-97(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Berkley;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blaise R.G., Champagne B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Gerhart W.M., Glaeser K.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei V., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relarte K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski R.P., Smith T.,
 RA Spier E., Spindling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitkas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng X.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RA Science 287:2185-2195(2000).
 RN [6]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkley;
 RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter C., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RA Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC STRAIN=Berkley; TISSUE=Head;
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.W.,
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.A.,
 RA Gonzalez M., Guatin H., Li P.W., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Paclab J.M., Paragas V., Park S., Phouanavong S.,
 RA Wan K.H., Yu C., Lewis S.E., Rubin G.M., Celinker S.E.;
 RA Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 RN [8]
 RP SIMILARITY TO C. ELEGANS CEH-26.
 RA MEDLINE=94212446; PubMed=7909177;
 RA Buerklin T.R.;
 RT "A Caenorhabditis elegans prospero homologue defines a novel domain.";
 RT Trends Biochem. Sci. 19:70-71(1994).
 RL -!- FUNCTION: Required for proper neuronal differentiation of most or
 CC all neurons and their precursors in central and peripheral nervous
 CC systems, axonal outgrowth and pathfinding. Not required for the
 CC specification of neuronal identity. May regulate transcription by
 CC binding to DNA.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=C; Synonyms=L;
 CC IsoId=P29617-1; Sequence=Displayed;
 CC Name=A;
 CC IsoId=P29617-2; Sequence=VSP_002306, VSP_002307, VSP_002308;
 CC Name=B;
 CC IsoId=P29617-3; Sequence=VSP_002307, VSP_002309;
 CC Name=S;
 CC IsoId=P29617-4; Sequence=VSP_002308;
 CC -!- TISSUE SPECIFICITY: NEURONAL PRECURSORS, EXPRESSED IN THE
 CC DEVELOPING CNS. LENS-SECRETING CONE CELLS OF THE EYE AND MIDGUT.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING
 CC FORMATION.
 CC -!- SIMILARITY: BELONGS TO THE PROSpero HOMEOBOX FAMILY.
 CC -!- CAUTION: Ref.7 sequence differs from that shown due to a
 CC frameshift in position 1122.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M81389; AAA28841.1; -
 CC EMBL; D10609; BAA01464.1; -
 CC EMBL; Z11743; CAA7802.1; -
 CC EMBL; AF190403; AAF05703.1; -
 CC EMBL; AE003691; AAF54628.2; -
 CC EMBL; AE003691; AAN13500.1; -
 CC EMBL; AE003691; AAN13501.1; -
 CC EMBL; AY060680; AAL28228.1; ALT_FRAME.
 CC PIR; S24548; S24548.
 CC PD3; IMIT; 04-DEC-02.
 CC FlyBase; FBgn0004595; pros.
 CC GO; GO:0045179; C:apical cortex, IDA.
 CC GO; GO:0045180; C:basal cortex, IDA.
 CC GO; GO:0005634; C:nucleus, IEP.

OY		394	-AEVDNSAETPRUYPAEFAAOULRKLSHSEHGSLGEBDVDRGS-----	436
Dd		407	GSDMDISLAS----PSHSZMDMLDKDVLDEDDDDDCVEQKTSGSGCKKREPMIDKARVE	462
OY		437	---PKQGRHPACSNANAGAPASIPRLAVNLHTMTMAAGIGAMPKLDETQTGVDFIKG	493
Dd		463	NIVSGNRCSBSSSLAOGQ-----LQVNGCKKKRKLQPORQHMER-----	502
OY		494	LIVANSGGIINEGL-LNILLSAQENSNGN--ASILLQQO-----OHQNHQHHQOOQ	543
Dd		503	--YVAAGAAG--LNFLANIQSWMILQDEDSSESNELESPOIQKRVEXNALKSQJRSMEQLAEM	568
OY		544	QOOHVAAHYAKRYLRPKSTPEPTNSSLD-----PNDASEPPLIKTISFKYSGERPASSSISP	597
Dd		561	QOKYVOLCSRMEESECDELDDDOVEOEBOEDMSSDHIELSPSLTLTGQG-----	612
OY		598	GLVGGHHNPLNNNSLSISNNNSHNSGNHRGNSNPSPHASPMV-AAAVALQGYVSAGNSL	656
Dd		613	-----DVSPIHKETGOBRPGSS--SP-SPEPLPKTSLGESSDSGANML	654
OY		657	LTSSSSSIQMMASNIOROIINEQSGESIRNG-----NVSDCSNNGSSSLGKKPSI	710
Dd		655	-----SQMSGSMMSGKLIHNPL-VGVGHRALPGCFRPLLQHMGMDSHAAMAYQOFFFEDEA-	708
OY		711	SVAKITGTDTSRFCGASPNNLLSQOHNSAHNLTHQOOOQSLSAQBALKGTRPKRGKYRNY	770
Dd		709	RMAK-----EAABEQOOOQOOOQOOOQOOOQEQRFRFEOEOORRRKEEQOOOI	758
OY		771	DRSLVEAKYAVORGEM-----SVHRAGSTYYCP-----HSLLEKVVER-	810
Dd		759	QROQO--OH!QO!QOQOQOQOEHAVATRAPRQMHNPAPRLPTRMGGAAGHTALSELSEKF	816
OY		811	HLMRPKRREPPO--PDLVGL-----	830
Dd		817	QMIRANNNSNMARMSTDLLEGADLVKSEITTSLSALVTITVFHFVQNRILFSKOADSVT	876
OY		831	GRANKLOLDKLYAG--PHGSKLSNALKNONNOAAAAAATAAATAAATPNGLKLPFE	887
Dd		877	AAAEOLNKDLLASQIILDRKSPRTKVADRPONGPTPATOSAAAFQAPKTPOGMN-FVAA	935
OY		888	A-----GP-----QALSFOPMFWOTNTATNYUGDPFRNITEMANPQ	925
Dd		936	AALYNMTGFCLPDPQOQOQOQTAAQOOSAQQOQ--OSSQOQTQOQLONEALSUVTPK	922
OY		926	ASNHGIMKSAODMVEN-----VUDGIIRKT-----LQ	953
Dd		993	KKH-----KUTDRIRIPRTVRSILLAODEVVRPTGSRPSTPQQOQOQOQOQOQOQOQOQOQOQ	104
OY		954	ASEGNGSAAANGSNSNGNGHGHGHGHGHALLDOLLVKYTPL-----PF	977
Dd		1049	ASNGGNSNATPAQSPTRSSCGAAYURPPRRPPMPVSLPTSVALIPNPSIHESKVSPY	110
OY		998	TNRHNDYATCSNAS-----GEVYKSGSGMGYADIKERLSADSGGSDE	104
Dd		1109	SPEFNPAHQAAGQATAQILOHNOHONRHNGHQSOLSSPPSGLAGL-----MDSRDSRPL	116
OY		1046	EHSASHINNNSDILANKKXSGGGGGGGGNGOT-----NENGSSPMWTSRD--DSETD	109
Dd		1163	PHEPSPHILPALAAAH-----GGSBDYKTCILRAYMDAQDQOSECNADMOFDQDMAR	121
OY		1097	ASSFFKSGENGCOONHKMDLNDLGSSSSSHIK-CESBAATGHNPSGHHTTSLHEKLAQI	1154
Dd		1215	TISF-----YKQWLKTEHESLMAHGCS--LTRPHS--STLRPMHLRKAKL	1258
RESULT 15				
CYAI DROME				
ID_CYAI DROME	STANDARD;	PRT;	2248 AA.	
AC P32870:				
DT 01-OCT-1993 (Rel. 27,				
DT 01-OCT-1993 (Rel. 27,				
DT 15-SEP-2003 (Rel. 42,				
DE Ca(2+)/calmodulin-responsive adenylate cyclase (EC 4.6.1.1) (ATP				

DE pyrophosphate-lyase) (rutabaga protein).
GN RUT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canon-S; TISSUE=Head;
RX MEDLINE=92154664; PubMed=173965;
RA Levin L.R.; Han P.-L.; Hwang P.M.; Feinstein P.G.; Davis R.L.;
RA Reed R.R.;
RT "The Drosophila learning and memory gene rutabaga encodes a
RU Cell 68:479-489(1992).
CC -1- FUNCTION: This is a membrane-bound, calmodulin-sensitive adenylyl
CC cyclase. Inactivation of this cyclase leads to a learning and
CC memory defect.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- COFACTOR: Binds 2 magnesium ions per subunit (by similarity).
CC -1- ENZYME REGULATION: Activated by calcium/calmodulin and G protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: MUSHROOM BODIES OF THE FLY BRAIN.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
CC cyclase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M81887; AAA28844.1; -
DR PIR: D42088; D42088.
DR HSSP: P19754; 1AMK.
DR Flybase: FBgn003301, rut.
DR GO: GO:0019933; P:camp-mediated signaling; NAS.
DR GO: GO:0007625; P:circadian behavior; NAS.
DR GO: GO:0007591; P:molting cycle (senu Insecta); IGI.
DR GO: GO:0008355; P:olfactory learning; NAS.
DR GO: GO:0045473; P:response to ethanol (senu Insecta); NAS.
DR InterPro: IPR01054; G_cyclase.
DR Pfam: PF00211; guanylate_cyc; 2.
DR SMART: SM00044; CYC2.
DR PROSITE: PS00452; GUANYLATE CYCLASES_1; 2.
DR PROSITE: PS01125; GUANYLATE CYCLASES_2; 2.
DR Lyase; CAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;
KW Metal-binding; Magnesium.
FT DOMAIN 1 41 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 42 60 POTENTIAL.
FT TRANSMEM 65 84 POTENTIAL.
FT TRANSMEM 101 115 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 152 174 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT DOMAIN 207 705 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 706 726 POTENTIAL.
FT TRANSMEM 730 750 POTENTIAL.
FT TRANSMEM 770 791 POTENTIAL.
FT DOMAIN 792 813 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 814 834 POTENTIAL.
FT TRANSMEM 842 867 POTENTIAL.
FT TRANSMEM 868 888 POTENTIAL.
FT DOMAIN 889 2248 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 318 334 CATALYTIC (POTENTIAL).
FT DOMAIN 1013 1029 CATALYTIC (POTENTIAL).
FT DOMAIN 515 530 GLY-RICH.
FT DOMAIN 569 602 GLY-RICH.
FT DOMAIN 1278 1297 GLN-RICH.

FT DOMAIN 1767 1810 GLY/SER-RICH.
FT TRANSMEM 2025 2040 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 2200 2241 GLN-RICH.
FT METAL 280 280 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT METAL 281 281 MAGNESIUM 2 (VIA CARBOXYL OXYGEN) (BY
FT METAL 281 281 SIMILARITY).
FT METAL 324 324 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT CARBOHYD 800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MUTAGEN 1026 1026 G->R; ABOLISHES CATALYTIC ACTIVITY.
SQ SEQUENCE 2248 AA; 248899 MW; E459C718BE01868 CRC64;
Query Match 3.9%; Score 237; DB 1; Length 2248;
Best local similarity 18.5%; Pred. No. 0.00066;
Matches 252; Conservative 126; Mismatches 382; Indels 604; Gaps 61;
QY 17 CMGRQWKHYQDKLT---C---SHLNIEEOPIALAGEDEPQYNSKREISQSNPNH 69
DB 1134 CRGTIKVKGKDMVTYFLCDSGKNSKNGEVNAMSLPQSLHAPDY---MKVSGFPENR 1189
QY 70 CKTENH-RLBOOH---NGSQLREED--SENQTSKSSKRTPTGATSTSPPEPIDW 122
DB 1190 VNTDTYSKKEGHHLYAGNG---VEEQQLLQHOKHND---PLP---LPAPP-PV-- 1235
QY 123 RPSAKCNFCVNGRLTLTVAQGLVAESAATATSSSTSHIHQSDSNSS--ASLPNHI 180
DB 1236 ---HHHLHQQOQQLNSKLOKQPIFM 1258
QY 181 SSSSSSN--NNSGCRARHIAASARATPAATPANSLEYKLTORAKMTSMDSMAQ 238
DB 1259 ANGGLPNIRENGNGHNSH---OQOQO 1282
QY 239 LAQFSLADFNILNSLASQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 286
DB 1283 QOQHQ-----OQOQHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1330
QY 287 KDTSPSPVDAP-----LDLSSKPSNNSISGVKVRACATPPSGRAVSEE-- 334
DB 1331 HPLPS-AVSVPVQHQLLHQLLOLOHQPVSVM---RENNIENPISGGRHOQEQP 1385
QY 335 ---DLRALQDVANKLDARKSAQNEBORSLDNRLFKMKIHDOEDHDELEDS 388
DB 1386 PHSGLSDLSGNGMGGA-----GVLGDDCFMPPRRDER----- 1419
QY 389 NDDEAEVDSNASTPVYPAEFAAQLKLSLSEHNSDLDEVDGSPKMGRRP---- 443
DB 1420 ---TYVPP-----LNQH-----GHNPPHHH 1437
QY 444 ACGNANAGAPASIPPL-----DANVLTHTLMLAAGIGAMPKLDFTQVDFIQLVA 497
DB 1438 SNLUNOSQHPSPFTSLGYGCCRSEBPLH---ASSVAPAKTIPMOHAPKY----- 1486
QY 498 NSGIMNEGLNLLSASQENSGNASLLQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 549
DB 1487 -----EPPRYTSPHTMLSQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1528
QY 550 AYRRLPK-----SETPTNSSLDPNDAE---DPILKIP-----SPFVS 586
DB 1529 QDPHPLQRYAMYSQQPQ---LPPKPVLRITMKPLPLPDLLESRDMSSTDLSRPH 1584
QY 587 GPASSS-----LSPGLVGGH--HH-----PL----- 607
DB 1585 SPSNSSSDESYSKTTGEGDEDESPKRVNGSHLHNRNGYLLPACGLVNPLOMLYPCIO 1644
QY 608 -----NNNSLSISNNSHNSHRNGSNPSHSPMLAAVAQ 648
DB 1645 VDPSPVDMALHDFELSTTEGQHTNTNTNTNTQKXDCSNGPFGK-----ATV 1698
QY 649 GYSAGNSLLTSS--SSSIOQKM-----ASNTIQOINDESG 681
DB 1699 GTAAGAAIATKSPFERELQRLINSSRARCLATATTGAIISTDTQTSNRSRELS----- 1754
QY 682 QESLRNGVSDCSSNN-CGSSSLGYKKRPSISVAKIIGTDTSR-----FGASPNLLSQOH 736

```
Db 1755 -YSLNGKLSANGVGGSSGSSGSSAANGSGGSSNGVLGGSSGSSNSGNN 1813
QY 737 SAHLITHQOQ---QQQLSAQELAKG-----TPKRGKRYNYRDSLIV 776
Db 1814 SSHKTEQQQNDHDLHLAGSKLLGSSNFWIAKHPVGLAIKEITRNKPSSESSQMTSDT 1873
QY 777 EAVKAV--QRGEMV-----HRAGSYG--VPHSTLEKVKER--HLMRPKREPK--PQP 824
Db 1874 ESCETLHNRNOMVLAIMEHTAKELNGSHAHGQHQQPQRTHQRPKSKELQYSHE 1933
QY 825 DLVGLTGPANKQLDLKLYAGPHGSKLSNALKNQNNQAAAAAATPGLKLP 884
Db 1934 SLDGLDG-----AVQSQSQ-----1948
QY 885 LFEAGPOLSPQNNFWPQTNTATNAYGDFNRITAMRNPOASNNHGLMKSQDMVENY 944
Db 1949 -----RNQRYHHHHHQROOQOQRYNHVQEOERDPTEDNLA 1986
QY 945 D-----GIIRKTLQASEGNSAANGNSGNSNGHGHGHALLDQLLVKKT 994
Db 1987 DEEFDEDEYGRDVRQKRLQKSE-----2008
QY 995 LPFTNRNDVYATCSASGESVKRSGSPMGNYADIKERLSADSGSDEEHSASHINN 1054
Db 2009 ---LNHKRSEVA-----TEAGNHDEVE--EEDDDDEEDHR-----2042
QY 1055 NNSDLAHNKNKSGGGGGGQNGNGRSGSRMTSRDSETPASSFKSGENGQONHKM 1114
Db 2043 -----NGGREAAPLTNGSMRGLNANVIND-----ELKYGAT--HLMHOSM 2080
QY 1115 DLNGSSSS--SHIKSEFAATGHS-----PGHHTTSLHE 1149
Db 2081 DSNPLESOSSEWSDDCREBATEGASTGYTTDEPLENISLINE 2124
```

Search completed: October 29, 2003, 12:28:18
Job time : 40 secs



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 107007

TO: Minh-Tam Davis
Location: cm1/8a01/8e12
Art Unit : 1642
Wednesday, October 29, 2003

Case Serial Number: 10/016768

From : Susan Hanley
Location: Biotech-Chem Library
CM1 6B05
Phone: 305-4053

susan.hanley@uspto.gov

Search Notes

RUSH

pty date:
10/2000

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:25:06 ; Search time 126 Seconds
(without alignments)
2385.963 Million cell updates/sec

Title: US-10-016-768a-10

Perfect score: 6030

Sequence: 1 MHISSEYISLSEKVAEECMGR.....ILHEKLAQIKAEQVDAQDL 1165

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_23.*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rident:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6014	99.7	1165	5 Q9VD60	Q9VD60 drosophila
2	4839	80.2	1221	5 Q24079	Q24079 drosophila
3	707.5	11.7	1598	5 Q95YM8	Q95YM8 apis mellif
4	336	5.6	2472	5 Q8T2M5	Q8T2M5 dictyosteli
5	329.5	5.5	1918	5 Q8SSW3	Q8SSW3 dictyosteli
6	314.5	5.2	2151	5 Q9VPL6	Q9VPL6 drosophila
7	314.5	5.2	5322	5 Q9VPL9	Q9VPL9 drosophila
8	311.5	5.2	2112	5 Q9VEL9	Q9VEL9 drosophila
9	310.5	5.1	2038	5 Q9W1L3	Q9W1L3 drosophila
10	309.5	5.1	2151	5 Q9Y1L3	Q9Y1L3 drosophila
11	308.5	5.1	1103	5 Q9VY72	Q9VY72 drosophila
12	305	5.1	1444	5 Q9VTN2	Q9VTN2 drosophila
13	305	5.1	1514	5 Q8S1S5	Q8S1S5 drosophila
14	304	5.0	1249	5 Q8INC6	Q8INC6 drosophila
15	302	5.0	1140	5 Q8T1T0	Q8T1T0 dictyosteli
16	302	5.0	1565	5 Q8T1M2	Q8T1M2 dictyosteli

17	296	4.9	1162	5 Q9W3Q5	Q9W3Q5 drosophila
18	295.5	4.9	1408	5 Q24341	Q24341 drosophila
19	293	4.9	1556	5 Q9VX23	Q9VX23 drosophila
20	293	4.9	3381	2 Q9KX33	Q9KX33 streptococ
21	293	4.9	4001	5 Q8WR07	Q8WR07 drosophila
22	292	4.8	2310	5 Q9GRA9	Q9GRA9 drosophila
23	290	4.8	1880	5 Q8MP27	Q8MP27 dictyosteli
24	287	4.8	1404	5 Q8IRY3	Q8IRY3 drosophila
25	285.5	4.7	939	5 Q9NH00	Q9NH00 drosophila
26	285.5	4.7	1409	5 Q9YX60	Q9YX60 drosophila
27	285	4.7	1721	5 Q8SSQ0	Q8SSQ0 dictyosteli
28	283.5	4.7	1180	5 Q24163	Q24163 drosophila
29	283	4.7	1416	5 Q9VYMI	Q9VYMI drosophila
30	282.5	4.7	1741	5 Q9W517	Q9W517 drosophila
31	282.5	4.7	1768	5 Q24153	Q24153 drosophila
32	281	4.7	929	5 Q9HBM0	Q9HBM0 drosophila
33	281	4.7	2280	5 Q9VBE6	Q9VBE6 drosophila
34	281	4.7	2302	5 Q9N693	Q9N693 drosophila
35	280.5	4.7	1741	5 Q46095	Q46095 drosophila
36	280	4.6	929	5 Q9HBL3	Q9HBL3 drosophila
37	279.5	4.6	929	5 Q9NGW5	Q9NGW5 drosophila
38	279.5	4.6	1186	5 Q9VED3	Q9VED3 drosophila
39	279	4.6	2150	5 Q23863	Q23863 dictyosteli
40	279	4.6	2181	5 Q9VRA6	Q9VRA6 drosophila
41	278.5	4.6	1673	5 Q9VQ19	Q9VQ19 drosophila
42	278.5	4.6	1785	5 Q81520	Q81520 plasmodium
43	277.5	4.6	1329	5 Q9W283	Q9W283 drosophila
44	277.5	4.6	2283	2 Q8VQ99	Q8VQ99 staphylococ
45	274	4.5	1338	5 Q8MN23	Q8MN23 dictyosteli

ALIGNMENTS

RESULT 1
Q9VD60 PRELIMINARY; PRT; 1165 AA.
ID Q9VD60
AC Q9VD60
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-OCT-2002 (TRMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE CG18389 protein.
GN EIP93F OR CG18389.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan M.R., Bouck J., Brockstein P., Broctier P.,
RA Borokova D., Borchan M.R., Bouck J., Brockstein P., Broctier P.,
RA Burris K.C., Busam D.A., Butler H., Cadelu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Fouts D., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Krat C., Kravitz S., Kulp D., Lai Z.,

[illegible]

[illegible][illegible]

Db 1463 OQQAQNEAAOGLAAAIASASTGGSGAGGKAAASASOLAAGFPFLPSPSLYPPWGLGGL 1522
 Qy 761 RP-----KKGXYRNYDRDLSLEAVKAVORGEMSVHAGSYVGVPHSTLEKVKYERHLMRPR 816
 Db 1523 NPYSIGSGGSLGSAVYQ--LAQOYNLINLNGATTSASTSTQSKSHOSKSSOSKSRRTTASA 1580
 Qy 817 KREPKRPPDVLGTGPANKLQDLKAGPHGGSKLSNMLKNQNNQAAAAAAMAAAAA 876
 Db 1581 NSASILMMAMAMSGCASTVTPSTASGSGRGSS--RNG-SCOTTTAAADLSSLL 1638
 Qy 877 TPNGKLPLFEAGP---QALSFPQWFWPQ-TNATNAYGL-----DNRITAMR 922
 Db 1639 MPG-----ADPHLESLSMSNMDLAQATRLMSLSGMPPLSGTSPSSGCGNSTSKR 1690
 Qy 923 NPOASNNHGLM-KSQADWVENYDGLIRKTLQASGNGSAAAGNGSNGNGHGHGHG 981
 Db 1691 SSQAANANVAQKEQKKWLESILARGALPTDLAALQAFSGCKMPSTSGN----- 1739
 Qy 982 HALLDQLVKTPTPLPFTNHRNDVATCSASGESVKSQS--PMGNVADIKERLSADS 1039
 Db 1740 -----TGSTSSSKSKAAATAAAQLPPIPMSSDFPQAFLEMAAQAAMA-A 1787
 Qy 1040 GGSDESHASHINNNSDLAHNNKXSGGGGGGNGGTNGGRSSRMTSRDDSETDASS 1099
 Db 1788 GGSILPLSGPGS-----LASLAGLTGSGAGGCG--SSASGSGTSHSSSKRQREDA-- 1834
 Qy 1100 FKSGNGGQONH--KMDLN-----CGSSSSSHIKESAAALGHSPGHHTTSI 1146
 Db 1835 FKQ-----QMDYLYTKTLGSGCISLIPTSAGSGSSASSAANAAAAA-- 1878
 Qy 1147 LHEKLAQIKAEVDQADQ 1164
 Db 1879 -----AALDAEQHQOQ 1891

RESULT 7

Q9VPL9 PRELIMINARY: PRT: 5322 AA.
 ID Q9VPL9 09VPL9: 09N164;
 AC 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE CG3696 protein (KISMET-L long isoform).
 GN KIS OR CG3660 OR CG3696 OR CG38326.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blasej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA LaSko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McInerison D.,
 RA Mervin G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Hock E., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mettel B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick P., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Therrien M., Morrison D.K., Wong A.M., Rubin G.M.;
 RT "A Genetic Screen for Modifiers of a KSR-Dependent Rough Eye Phenotype
 in Drosophila.";
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE003590; AAF5157.3;
 DR EMBL: AF215703; AAF43004.1;
 DR HSSP: P23197; IAP0.
 DR FlyBase: FBGN001309; kis.
 DR InterPro: IPR000953; Chromo.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N_
 DR Pfam: PF00271; helicase_C_1.
 DR Pfam: PF00217; SNF2_N_1.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.

OY	96	OTSHSRTPTPGASTPSTPRPEPILDMRPSACNFCVNRLLITVAOGLVESAATTS	155
Db	3847	OTSHHQVYASSASASTSGVDVOTQSKRS-----SLSRLGLPAKSTINAAVAAVAAA	3900
OY	156	SSTSNSHITHOHSDSNSASLPHH1SSSSSSNNNSG-----NRARHIA-	199
Db	3901	AAGN-----ATSLSSLLPGMSLAAATSSSAGVSGLSVPASGVGKKRRHIAI	3955
OY	200	-AASARATPAAATPANSLLEYLLTORAAKMTSMDSMAQLAOFSLADFNL-	INSLAS 256
Db	3952	DVETERA-----KLHALLNSSTMA PKDWESEIANMEALGSGSGRGGSNASS	4000
OY	257	OOOQOOOOOOLASAVPTTSEVAAA1SPALKTPEPSVDAPLDLSK-----	PSB 306
Db	4001	GMOPPPAHOHASLSROSSGOFKSPAV-PAMK-TPEPSSGAPWDLSSSLPKMMTEMLXA	4056
OY	307	NSISIGDVKSYACATPT-----PS-GRRAYSEEDLSRALODVANKLDARK	352
Db	4059	SSSGAIDLSEVODPFMPKSKSVHAALSAFFPSMKNKSLDITLKMK--KNCTIE	4111
OY	353	SASQNHQPSILDNLFPMKHHHDOBDHGDLEDSNDDAEYVDSNMS-TPVPAEFAR	411
Db	4116	PVIGEKERKKKXDEIVLGLSAKKQKTFPPDPLPSKK--KQIIPPSVTP-----	416
OY	412	AQLRLSHLSEHN-----GSDLGEDVDKSPKGRHPACGNMSANGABA----	456
Db	4155	ANVGSSNQOQSKQKFTTIVTTVPKSKSGSSNSGSGTGSSASAGGSGGLSALOM	422
OY	457	-----SIPLDANVL-----HTLMLAAGIGAMPKX-	481
Db	4225	AMGSLSKDLSMALLAQTMAITPOTFLKQOQGMQFLPQAOKKAYENMLAEHQAM-KIS	428
OY	482	-----DETQVGDFTIKGLLVANSGGITMEGLINLLASQOENSNGNS	523
Db	4284	SKFSTNSPHDVKNKMLSDMTSPLDQOL-SIDYVGSGSGSGSGNSRKSNNQOQNSQS	4344
OY	524	LLLOOQOHOQHOOHOOOQOQOQVAAVYHRLPKSETP-----ETNSSLDPNDA-----	572
Db	4333	SSAAOLOKQOQOQOQOQOQOQOQSHSPGPNLTGEEBPVPTNKQTKRIGCNPAPOLKRLM	440
OY	573	---SEDPILKI-PSF-KVSGPASSSLSPGGLVGH-----HPLNNNSLSTSN	617
Db	4403	QWLTENPYVEVDPKMLEQOMQNMSTPSPRPASMESGYSSAKVSHOGRPLSLNSTSSSS	446
OY	618	NSNHSNSHNRN--GSN-----RSPHSGFMLAAVAAGVYSAQ-----	653
Db	4463	HTQOQSSAAOQAGNCGSSKKKNSRQQTAAASALDQALLOFSLGLNPSILANTPLGLGA	4522
OY	654	-----NSLITSSSSSIOKMAASNORQINBSGQESLRNGVSPCCSSNNGS	700
Db	4533	FDPKNPPLAFDPRKNPLSWSFGGMPDM-GN-----PGJGNLNNMLFASLGMGGL	4573
OY	701	SSL-GYKKPSI-----SVAKITIGDTISRFGA-----SPNLSQOHNASHNLTHOO	746
Db	4574	CNLAAMDPTQSLAALMAAGPTLGLTGAAGCAGSGKSNQOSQSSATSSSSSKKKKQOQ	4633
OY	747	OOQLSAQEAL-----GK-----GT	760
Db	4634	OOQAOMEAAOLAAASISTGSGAGGKNAALASASQLAAGPFLFPNPFLYRPMGLGL	4693
OY	761	RP-----KXGKYNNYDRDLSVEAVKAVORREMSVNHASGYUYVGRHSTLEKKUVERHLMR	816
Db	4634	NPYSLGSSGSLGAYAO--LDAQYNNLNGATSSASNTSTQOSKHOSQOSKQSSRNTTASA	4755

RA Nelson D.R., Nelson K.A., Nixon K., Nuss

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

Query Match	5.2%;	Score 311.5;	DB 5;	Length 2112;
Best Local Similarity	18.9%;	Pred. No. 5.6e-09;		
Matches 225; Conservative	167;	Mismatches 411;	Indels 389;	Gaps 41

Db 1069 SSSSSSSNGSSSNTGSSSSSSNGASSSSGGSSNQSSSSSSNGSSSSGSSSSGSSSSGSSNQSTS 1120

RA Laasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei J., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milentina N.V., Moatry C., Morris J., Mostreli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Pitzman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Wooddey T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 [12]
 RP SEQUENCE FROM N.A.
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreeneck D., Farfan D.,
 RA Ferrera S., Fris E., Galle R.F., Gang N.S., George R.A.,
 RA Gonzalez M., Hout J., Hoskins R.A., Hostin D., Howard T.J.,
 RA Ibsenham C., Jallai M., Kruse D., Li P., Mattei B., Mostreli A.,
 RA Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pitzman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [13]
 RP SEQUENCE FROM N.A.
 RA Mira S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Ceiniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Fris E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Asburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [14]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [15]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 DL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE003442; AAF46312.3; -
 DR HSSP; 092831; 1891.
 DR FlyBase; PBgn0004656; fg(1)h.
 DR InterPro; IPR00104; Antifreeze_1.
 DR InterPro; IPR001487; Bromodomaim.
 DR InterPro; IPR002173; PfKB.
 DR Pfam; PF00439; bromodomaim; 2.
 DR PRINTS; PR00308; ANTIREEZER.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SMO0297; BROMO; 2.
 DR PROSITE; PS00633; BROMODOMAIN_1; 2.
 DR PROSITE; PS00104; BROMODOMAIN_2; 1.
 DR PROSITE; PS00583; PFKB KINASES_1; 1.
 SO SEQUENCE 2038 AA; 205346 MW; DC4A1A7B1266191E CRC64;

D	b		1064	SASSVEVAVQANPVSSSSSSS-----DSSSSSSSD-----SSSSDSDSFAGGODE----	1110
Q	y		200	AASARATPFAATPANSELELYKLTORAAKMTSMDSMAAQLOFSLADFNLI--NSLASQ	257
D	b		1111	-----RPRKKSKRSDSGSNVNPSI-----NVVMGNLPS-	1141
Q	y		258	QQOQQOQQOLASAVPTT-----SEVSAAALSPALKOTPEPSVDAPLDLSKPe-P	306
D	b		1142	-----GALSPTTMLMGLDHVNSNTPTSQISMNLGANPLTLAAMLNNNKTSLP	1191
Q	y		307	NSSISGDVKYACATPPPSGRRAAYSEEDLSRALODV-VANKLDARKSAQHNEQRSLTD	365
D	b		1192	GSNREG-----AAPRG-----NMNHAAGVPVNAQAASASCGQHN-----	1227
Q	y		366	NRLFRMKHDEQOHDDGEDESDNDDAEAEVDSNASPTVPYAERF-----	410
D	b		1228	-----KNGPNDSLKYOPGGPI-----NAALP--PHSFAGGTATVATSQSGGI	1268
Q	y		411	--PAQLRLSLSHENGSDGED-----VDRGSPKMGRRPARCAGMASANGAPASI-PL	460
D	b		1269	RIASLHLRPSGL-----GGDDLHEHAALAALTGINSTG--TAAGGINNGSGNNANPL	1322
Q	y		461	-----DANVLHTLMLAAGIAMPKLDETQ-----TVGPETIKGLLVANSIGINEG	506
D	b		1324	GSMSHDAMVNAVSLASLSGLKOITQFPDDPVESQSLASFASGSRGKSGLTGNF--LMQNH	1381
Q	y		507	LNLILSASQENSNGVASLLLOQOO--HQNHQHHQOQOQOQVAAYRHRLPKSETPET	563
D	b		1382	LMOPAFQOQOQOQ-----OQOQPFGHQOQOQOQOQOQOQOQHM DY-----	1422
Q	y		564	NSLDPDNDASEDLIKLPFKVSCPASSSSLSPGLVGGHHHPMLNNSLSISNNSNH--	621
D	b		1423	-----VTBLLSIGAVNWGMNG--NHLINFLMDMAAAVOOKRPQ	1459
Q	y		622	--SSNSHRNGSNSRSPHSAPMLAAVAOCGVASGNSLTTSSSSI-----QJMMASN	671
D	b		1460	OQOQOAHNNGN-----VADFCAAGFDGLMTATASFLDLEPSLOQOQMOMQ	1506
Q	y		672	IQRQINESGOESLRNCNVSDCSSNNGSSSLGYKKPISIVAKIIIGTDTSRFGASPML	731
D	b		1507	LQOQHQQOQOQOQT HQO-----	1522
Q	y		732	SOQHSHAHLLTHQOQOQOOLSAGOEALGKTRPKRKRYRYPBDLSLVEAKVQREMSVHR	791
D	b		1523	QOQHQQOQH--QOQOQOQOULTQOQ-----LQOQO-----	1548
Q	y		792	AGSYGVPHSTLEYKKEKHLMPRKE-----PKPOBDVLGLTPANKLOLD	839
D	b		1549	-----OQOQOQOQHLOQOQHQQOQHQAANKLLITKPRESM--MSPRDKOLO	1594
Q	y		840	K-LXAGPHGSKLSNALKNQNNQAAAAAAAATAATPNGLKLPFEAGPOLSPON	898
D	b		1595	QHOKVLPFOQS--PSDMKLHPNAAAAA AVASAQKLVYT-----FKANEQULKNASS	1644
Q	y		899	MFW-----PQTNATN-----AYGDFNRRTTE	919
D	b		1645	--WSSLASANSPOSTSSSSSSSSAXAKRAMDSFOOFNKAKERDYLKLEAEKEREKQKE	1702
Q	y		920	AMRNPOASHNHGLMKKSADMWEENVYDGIIRTTLOASE--GNGSAGNGSGNGNGCHGH	976
D	b		1703	AAEKEOQRKHKKSSSSSLTSTAVALOAAAIAAATAAVALTLAABAAAALASASNPSSGSS	1762
Q	y		977	GHGHHALLDOLLVKYPTLPFTNHRNDUYATCSASGESYKRSGSPMGYNADIKRERTLS	1036
D	b		1763	SGGAG-----STSGOALTGORDBR-----RDREBERER	1790
Q	y		1037	ADSGSSDEHSASHINNNSDLAHNKN-----KSGGGGGGGNGQTINGNCRSSRMPTS RD	1091
D	b		1791	SGSGGGGCGG-----GNNSSNSANSNGPSSAGSGGGGGGGGSPASAGCPNS-----	1837
Q	y		1092	DSEITDASSFKSGENGCQOHNHKMOMLNGSSSSSHIKCESEAATGHS	1138
D	b		1838	-GGGTATANSNGGGGGGGPAL--LNGSNSNGSVGGGAASNSNS	1861


```
RESULT 10
09Y1L3 PRELIMINARY; PRT; 2151 AA.
ID 09Y1L3
AC 09Y1L3;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE KISMET
GN KIS OR CG3660 OR CG3696 OR CG18326.
OS Drosophila melanogaster (Fruit Fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ISO-1;
RX MEDLINE=99146882; PubMed=10021337;
RA Dabreesse G., Deuring R., Moore L., Papoulias O., Zakrejsek I.,
RA Waldrup W.R., Scott M.P., Kennison J.A., Tankun J.W.;
RT "The Drosophila Kismet gene is related to chromatin-remodeling factors
RT and is required for both segmentation and segment identity."
RL Development 126:1175-1187(1999).
DR EMBL; AF11847; AAD37500.1; -.
DR FLYbase; FBgn0001309; kis.
DR InterPro; IPR006576; BRK.
DR SMART; SM00592; BRK; 1.
SQ SEQUENCE 2151 AA; 224802 MW; 021425B7729339CF CRC64;

Query Match 5.1%; Score 309.5; DB 5; Length 2151;
Best Local Similarity 20.7%; Pred No. 7.4e-09;
Matches 278; Conservative 167; Mismatches 497; Indels 401; Gaps 56;

QY 96 QTSHDSRTPTGATSTPSPPEPIDMRPSAKCNFCVNGRLTVNAGKLVESATATTS 155
DB 676 QTSHQVVASASAASTSGVPQVQSKPS-----SNSPLGLDKSINAAVAAVA 729
QY 156 SSTSNHITHQSDNSASLPHHISSSSSNNSSG-----NRAPHTA- 199
DB 730 AAGGN-----RTSLSLPLPGMSLSATGTSAGVSGLVSPSAGSGVCKRKHIAI 780
QY 200 -AASATPAAATPANSLELYLTLTQRAKMTSMDSMAQLAOFSLADPNL--INSTAS 256
DB 781 DVETERA-----KLHALNNSSTYAPKDWSEIANNMELSGSGRGKGNSSASS 829
QY 257 OOOOQOOQOQIASAVTPTTSEVSAALISPAKIDTPSPVDAPILDSK-----PSP 306
DB 830 GMPPRAHQHSLSSQSGQFSKPAV-PAMK--TPPSMGAPMDLSSSLPKMMTEMLKSA 887
QY 307 NSISGDKVSVACATPT-----PS-GRRAVSEEDLSFALDDVANKLIDARK 352
DB 888 SSSGAIIDISEVQDFSPWPKSSVHAALSAPFPMGKSKLDTLNLKMK---KNCTIEE 944
QY 353 SASQHNEQSLIDNLFKMKHHDOEDHDGDELDSDNDAEAEVDSNNS--TPVYPAEFAR 411
DB 945 PVIGKEKKKKKLDEIVLGLSAAKEQRTPPDPSPSK--KQOIPPSVVTTP----- 993
QY 412 AOLRLKLSLSEHN-----GSDLGEDVDRCSPKMGRIAPACGNASANGAPARA---- 456
DB 994 ANLGSSSSNQSQKQKPTTIVTTVPKSKSGSSSSSGTSGSSASGCGGGLSALQNM 1053
QY 457 -----SIPLDANVL-----HTMLAAGIGAMPKL- 481
DB 1054 AMGGLSSKDSLALLAQTWATDPOTFLKQOQKMMQFLPPAPORKAVERNLAEMQAM-KIS 1112
QY 482 -----DETQVGFPIKGLLVANSGGIMNGGLNLLSASENNGNNS 523
DB 1113 SKFSTNSPVDVYVVKMLSMTSPFLGDQL-SIDYVGSGGASGSGSRRSRNQOQGSQOS 1171
QY 524 LLLQOQONQONQONQOQOQOQO-----VAAVRRRL 555
```

```
DB 1172 SSAQLOKQO-QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1230
QY 556 PKSETPTNNSIDPN--DASEDPLIKTPFKVSGPSSSSLSPLGLVGH--HPLANNNS 612
DB 1231 MOWLTENPYEVDPPKLEOMQNP-NSTPSPR--PASMESAYSSAVSHGCRPLSNLSS 1286
QY 613 LSISSNSNSNSHN--GSN-----RSPSHASPMIAAIVAAGGYSAG----- 653
DB 1287 TSSSHHTQOQSSAASQAGNGSGSKNSROOTAASALDQALDQALDQALDQALDQALDQ 1346
QY 654 -----NSLITSSSSSIQXMMASNIQRIQNEQSGESLRNGVSDCS 695
DB 1347 PELGAFDPKPNPLAAPDKNPLILMSFGGPMQ--GNI-----PGLGNLNMNI-PASIA 1397
QY 696 NNGGSSSL-GYKKPFI-----SVALIGTDIRGCA-----SPNLSQHHSAHL 741
DB 1398 GNGGLGNLGNMDTOSLALMAAAGPTLGLTGASGAGSGKSQAOSQSSATSSSSASAK 1457
QY 742 THQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 758
DB 1458 KQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1517
QY 759 ---GTRP---KRGYRNYDRDLSVEAVKAVORGEMSVHAGSYGVPHSTLEYKVERH 811
DB 1518 GLGGLNPYSLGSSGLGSAVDQ--LAQOYNLNGATTSASNTSSTQSKHQSGSKSQSRN 1575
QY 812 LMRPKREPKEPQDVLGLTPRANKQLDKLXGPRGSKSLNALXVONNQAIAAAAAA 871
DB 1576 TTASANSASLNNANASMGSGASTVTPSTSSGSGRROSSES-RNQ--SQTTPAADVAO 1633
QY 872 AAAAATPNGKLPLFEAGP---QALSFPNMFWPO--TVNATNAYL-----DFNRI 917
DB 1634 LSSLMPG-----ADPHLIESLSRMSNMDLAATLMSGLNMPPLSGTSSGGCGNT 1685
QY 918 TEAMENPOASNHGLM-KSADPMVENYDGIIRKTLQASEGNSAAGNSGNSGNGHGH 976
DB 1686 STSKRSQAANEAVQAEOQKWLSELRGALPTDLALQAFSGQKMPSTSSN----- 1739
QY 977 GHGHALLDQLVKKTPPLPFNNRNDYAATCSSASGESVYKSGS--PMGNAYDIKER 1034
DB 1740 -----TGTSTSSSKSKSAATTAALQOIPQIPGMSDPPQAFIAMAQA 1783
QY 1035 LSADSGSSDEHSASHINNNSDLAHNKNKSGGGGGGQNTGNGRSMSTSRDSE 1094
DB 1784 MAA-AGSILPUSPGS-----LASLAGLTGSAGGG--SASGSTSHSSKRORE 1831
QY 1095 TDASSFKSGENGQGNH--KMWDLN-----GSSSSSHIKCESEATGHHSPGH 1141
DB 1832 QDA--FKQ-----QMDYTTKTLGLGSGSLIPTSSAGSSSASSAANAAAAVA----- 1878
QY 1142 HTTSLHEKLAQIKAEQVDADQ 1164
DB 1879 -----AALDAEQHOQOQ 1891

RESULT 11
09YV72 PRELIMINARY; PRT; 1103 AA.
ID 09YV72
AC 09YV72;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2003 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CG32611 protein.
GN CG32611 OR CG11075 OR CG11082.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
```

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Aamatiadis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
 RA Branton R.C., Rogers Y.H.C., Bialek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrell J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Bortman M.R., Bouck J., Brockstein P., Broctier P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervlov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sideris-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [12]
 RN SEQUENCE FROM N.A.
 RP Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Aamatiadis P.G., Branton R.C., Rogers Y.,
 RA Batzon J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galie R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RA "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [13]
 RN SEQUENCE FROM N.A.
 RP Mirra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Rupp J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RA "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [14]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [15]
 RN SEQUENCE FROM N.A.
 RP FlyBase;

RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 CC -! SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC EMBL: AE003494; AAF48333.2; -!
 DR FlyBase: FBgn0052611; CG32611.
 DR InterPro: IPR000104; Antifreeze_1.
 DR InterPro: IPR000210; BTB_POZ.
 DR InterPro: IPR007087; Znf_C2H2.
 DR InterPro: IPR007086; Znf_C2H2_sub.
 DR Pfam: PF00651; BTB: 1.
 DR Pfam: PF00096; ZF-C2H2_4.
 DR PRINTS: PF00308; ANTIFREEZE1.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR SMART: SM00225; BTB: 1.
 DR SMART: SM00355; ZNF_C2H2_4.
 DR PROSITE: PS00097; BTB: 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1_4.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2_4.
 KW Metal-binding; Nuclear protein; zinc; zinc-finger.
 SQ SEQUENCE 1103 AA; 113716 MW; EB586F5683C9B0B CRC64;
 Query Match 5.1%; Score 308.5; DB 5; Length 1103;
 Best Local Similarity 20.3%; Pred. No. 3.6e-097;
 Matches 221; Conservative 130; Mismatches 366; Indels 373; Gaps 40;
 QY 147 AESAATATSTSTSNHIOHDSNSASLPFHISSSSSNNNSGNRRARIAASARAT 206
 DB 120 ATQATATATATAA-----ATSTSTATSAATAATAATATATATATAT 160
 QY 207 PAATPANSELYKLTORAAKMTSMDSMAAQLAFSLADFNLSIASQQQQQQQQQI 266
 DB 161 ATAAAGSSN-----TTTTTAATTATCATTA-----T 187
 QY 267 ASATPTTSEVSAAAIAPALDTPSPVDAPLDLSKPSNPSSISGVKSVACATPTPS 326
 DB 188 STPATSSSSSVTSAAAAA---AAASAGVAHEEATSSSTSGCGKREASDSSPTPA 243
 QY 327 GRAYSEEDLSRALQDVVANKDARKSASQHNHQRSLDNLFPKHHHDEQDHGDELE 386
 DB 244 KRSSRMHP-----ADKVADVEDREQNQADLELSEELGRLKDEEDDVDEL 293
 QY 387 DS-----NDDEAEVDSNASTPVY-----PAEPAQQLRKLSHS----- 421
 DB 294 ENEETKLRGDEDDDEDEEDTPEPLDLYORPVEPKSAAATAAQQGSSNSISGSNA 353
 QY 422 -----EHNGSDLGEVDVGRSPMGHNPACGNASANOQAPASIPLD----- 461
 DB 354 CHTNSSNSNNNNNNNNNNSSNNNTSSGKTASNSGTAASGCTADSVAAVDDDDDD 413
 QY 462 -----ANVLHTLMLAAGICAMPKLDGTQVDFIKGLLVANSQGINNEGILNLSAS 514
 DB 414 DGGHHQKQVW-----DRLQLQVDVEEDLDDDDVVVPATAMARKIAQRL--A 459
 QY 515 QENSNGNASLLQQCQHQHQQHQQHQQCQQAAYARHRLPKSETPTNSSLPNDASE 574
 DB 460 HON-----LQRLHTTHHQAQ--QHSHHHPHSQHHHPHQHHTHS--DDEBAM- 506
 QY 575 DPLIKTSFKVSPASSSSSLPGLVGHHNPLNNNSLSISNSNHSNRGNSRSP 634
 DB 507 PVI-----AKSEILD-----DYDEMDLEDDDEADNSNDGLNKK- 543
 QY 635 HSASPMIAAVALAOGG--YSAGNSL-----TSSSSISIKMMASNIQROINQSG 681
 DB 544 -----MGSGAGGGGGVDTSTGLPLPLTPSSAAAAAAMAMEQRTSPMWSA 597
 QY 682 QESLRNGVSDCCSSNNNGSSSLGYKPKSISVAKIIGTDTSTRFAS-----PRL 731
 DB 598 AQAAGAG-ASDLSTGGCGG-----RPSRSSRDPGGGNDGSGASSSSLSPTGVNLT 650
 QY 732 -----SOCHSAHLLTHQQQQQQLASQAEALGKGRPKRGKVRNDRPSLV 776
 DB 651 PVQSVPLSLKKEIDSSDSSH--SRHMOORSASAGGGLG-GLDYPASHESSTSES-- 705
 QY 777 EAVRAVGEMSVARAGSYGVVPSTLEYKVKERHLNRPKRREPKPODVLGLTGPAK 836

OY		230	TSMQMAQLAOLPILADPNLINSLASOQQOOQQOQOQINSAVTPTSEVSAAISPALDPT	289
Dd		291	SSTEQSDPTTOESSSTEGLSTESPAVTDOSSSTESSODSTTOESSSSTEGPLSTES	350
OY		290	PSPVDAPDLDSKESPNSISGDVKSVACATPPSGRRAYSEEDLRALADVPANKLD	349
Dd		351	---STEATNESSSTESSODSTTOESSSSTEGPLSTESTEATNESSSTESSQD-----	400
OY		350	ARKASQHNHQRSILDNLFPKKKHHDDQEOHDHDELEDSNDDAEEVDSNASTPYRPAEF	409
Dd		401	----STTOESSSSTEGPL--STESSTEATNESSSTESSODSTTOESSSSTEGPLSTESS	453
OY		410	ARAQLRKLSHEHNGSDLDGEDVDGSGPRMGHRPACGNASANOQAPASIPDANVLHT-	468
Dd		454	TEATNESSSTESSODSTTOESSSSTESP-LSTEP---STEANESSSTESSODSTTOESSS	509
OY		469	----LMLAGIGAMPKLDETQVGDFIKGLLVANSGGIMNEGILLNLASQENSGNAS	523
Dd		510	STEDPLSTESSSTEATNESSSTESSQD-----STTOESSSSTEGPLS-TESSTEGENESS	563
OY		524	LLOOQOQHQQHQQHQQOQQOQOQHAAIRHLPKSETFETNSSLDPND-----AED	575
Dd		564	TESSDSTTOXSSSSTESPLSTE-----PSTEANESSSTESSODSTTOESSSSTEG	614
OY		576	PILKIPSEKVSAGPA-----SSLSLPGCLGVGHNNHNNN-----NSLSISNNS	619
Dd		615	PLSTEPSTEANESSSTESSODSTTOESSSSSBG-----PLSTESSTEANESSSTESS	666
OY		620	NHSHSNHGANGSRPSHSASPMLAAVAOGGYAGNSLLTSSSIQCKMASHIQOIQIEQ	679
Dd		667	QDSTTOESSSSTESPLETEPSTEANESSSTESSODSTTOESSSSTEGPLSTEPSTEANES	726
OY		680	SGOEELRGANGVDCSNNGSGSSLGYPKPISIIVAKIGTDPSRFAGSPNLSQOHGAH	739
Dd		727	SSTESSQDSTTOESSSSBG-----PLSTESTEANESSSTESSODSTTOESSS-	776
OY		740	HLTHOQOQOQSABALGKTRPKRGGKYRNDRDSLVEAVAVOREMSVBRAGSYGVF	799
Dd		777	--TEDPLSTESSTEATYESSSTESSODSTTOESSSSTEGPLSTESSTEGSNSSSTESSQ	834
OY		800	HSTLEYUKERHLMRPKRREPPOBDVLGLTERPAKKLDQDKAKAPHG-SKLNSALKNQ	858
Dd		835	DSTIOESSSSTE--SPLSTEPSTEANESSSTESSODSTTOESSSSTEGPLSTESSTEANE	892
OY		859	NNQAAAAAAAAAAAAATPNGLKLPLEAFGOALS-----OPNMFW	901
Dd		893	SSSTESSQDSTTOESSSSTEGPLSTESSTEGENESSSTESSQDSTTOESSSSTESPLSTE	952
OY		902	POTANTNAVGDLFNKITTEAMR-----NPQSNHHGLMKSAODVENVYDGII	948
Dd		953	PSTEANESSSTESSODSTTOESSSSTEGPLSTESSTEANESSSTESSQDST-----	1004
OY		949	RKTLOASGEANGAANGSGNGNGNHGHGHGHALLDOLLVAKTPLLPTNRNNDYAT	1008
Dd		1005	-----TOESSSSTEGPLSTESSTE-----VTOEPSPRESI.PNSSTOQT	1043
OY		1009	-CSASAGESVKRSGSPMGNYADIKERLSADSQSGSDEEHASH-----INNNSDLAHN	1067
Dd		1043	PCTTNPSPSLBEPSPPTGPN-----DDDBSGNAGEKNGSTGSPCTTNPNPDPEBS	1093
OY		1063	KNKSGGGGGGGGCGGTNGNGGRSSRMTRSIRDSTDASSPF-----SCENG-----	1100
Dd		1094	SSTPGNDDSGNSGENGSSTGSGPCITTDNPSIDPESSSSTPGANDDGSNCSGSEGITST	1155
OY		1107	-----GQONHKWMDLNGSSSSSHIKCESEA-----TGHSFP	1133
Dd		1154	TGAPYTDTNPAQDEBPSPAENPBGDSNGSSSPREGATPTCPNAPKKSSTSYTAHPFP	1211
OY		1140	GHHNT-----TSILHEKLAQIAAEYOQDAD	1163
Dd		1214	KYTEGKAETSTLKSPTGTGHOEDRD	1243

```

RESULT 13
08SY55
ID 08SY55 PRELIMINARY; PRT; 1514 AA.
AC 08SY55;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE GH09355P.
GN CG6004.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Paclebo J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisner S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY075323; AAL68190.1; -.
DR FlyBase; FBgn0036203; CG6004.
DR InterPro; IPR002557; Chitin_bind_perh.
DR Pfam; PF01607; CBM_14; 4.
DR SMART; SMO0494; ChEBD2; 4.
SQ SEQUENCE 1514 AA; 157140 MW; 1FQC4B064105ADS CRC64;

```

Db 634 TESSODSTTQKSSSTESPLSTE-----PSTEANSSSTESSQDSTTOESSSSTEG 684
 Qy 576 PILKIPSFVSGPAS-----SSLSIPGVLGVHHHPLNNN-----NSLSINNS 619
 Db 685 PLSTPSTPEANSSSTESSQDSTTOESSSSTEG-----PLSTESSTEANSSSTESS 736
 Qy 620 NHHNSHRGNSRSPHSPMLAAVAAGVAGSNLSLTSSSSSIQKMAWNIQOINQ 679
 Db 727 QDSTTOESSSSTESPLSTPEANSSSTESSQDSTTOESSSSTEGPLSTPEANSS 796
 Qy 680 SCQSLRNKNNVDCSSNNGSSSLGKYSIVAKTIGTDTGRCASPNLLSQCHSAH 739
 Db 797 SSTESSQDSTTOESSSSTEG-----PLSTESSTEANSSSTESSQDSTTOESSS 846
 Qy 740 HLTHQOQOQLSAQALGKTRPKGKXANYRDSLVEAVKAVQREMSVHAGSYGVP 799
 Db 847 --TEBPLSTESSTEATYESSSTESSQDSTTOESSSSTEGPLSTESSTESSSTESSQ 904
 Qy 800 HSTLEYKVKERHMPREKPEKPPDVLGTPANKLQDLKACPHG-SKLSNALKNQ 858
 Db 905 DSTTOESSSSTE--SPLSTPEANSSSTESSQDSTTOESSSSTEGPLSTESSSTEANE 962
 Qy 859 NNQALAAAAAATAATPNCUKLPFAGPOALSF-----QPNMF 901
 Db 963 SSTESSQDSTTOESSSSTEGPLSTESSSTESSSTESSQDSTTOESSSSTESPLSTE 1022
 Qy 902 PCTNATNAGLDNFRTEAMR-----NQASNNHGMKMSADWENYDGI 948
 Db 1023 PSTANESSSTESSQDSTTOESSSSTEGPLSTESSSTEANSSSTESSQDST----- 1074
 Qy 949 RKTLOASEGNGSAGNGSNGNGHGHGHALLDOLVKYKPLPTNNRNDYAT 1008
 Db 1075 -----TOESSSSTEGPLSTESSSTE-----VTQESPLSTESPLNSTOCT 1112
 Qy 1009 -CSSASGSEVKRSGSPMGYADIKRELSADSGGSEDEHSAH-----INNNSDLAHN 1062
 Db 1113 PCTTNPSLEPSPPTPGN-----DDSGNSGSENGNSSTGSPCTTNPSPESS 1163
 Qy 1063 KKKSGGGGGCGGNGGNGGSSRMTSRDSETDASSFK-----SGENG----- 1106
 Db 1164 STPGNDDSGNSGSENGNSSTGSPCTTNPSDPESSSTPGNDDSGNSGSEGTST 1223
 Qy 1107 -----GQONHKMMDLNGSSSSSHIKCESEAA-----TGHSFP 1139
 Db 1224 TGAPYTTNPAQDESPAPENPGDSGNSSPBGAITPCTPNAPKSTSTSYAHPT 1283
 Qy 1140 GHHT-----TSLHEKLAQIAEQVDQAD 1163
 Db 1284 KYTTEGNKATSTLKSPTGTTPGHQEDRTD 1313
 RESULT 14
 Q8INC6 PRELIMINARY; PRT; 1249 AA.
 ID Q8INC6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE CG392-PB.
 GN SRP.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_taxid=7227;
 OK NCBI
 RN (1)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazew R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson K.C., Gabor G.L.,
 RA Axel J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bergman B.P., Bhandari D., Bolshakov S.,
 RA Borrova D., Botchan M.R., Bouck J., Brostein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glaser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoston D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kratz C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter J., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amandides P.G., Brandon R.C., Rogers Y.,
 RA Barzon J., An H., Baldwin D., Barzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gang N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hoston D., Howland T.J.,
 RA Ibegwam C., Jalili M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tuzy J.L., Bergman C., Bernan B., Carlson J.W., Ceiniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn R., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RP Flybase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003711; AAN13691.1;
 SQ SEQUENCE 1249 AA; 133783 MW; 83CDAD995A7DB7DB CRC64;

Query Match 5.0%; Score 304; DB 5; Length 1249;
 Best Local Similarity 18.9%; Pred. No. 7.6e-09;
 Matches 243; Conservative 176; Mismatches 470; Indels 396; Gaps 49;

```

Oy 83 GSOLLEEDSENNOTSHDSRTTPPGATSTSPPEPIDMRPSAKCNFCVNGRLITVA- 141
Db 82 GSMAANAEETIAIETENGEAATPTAAATA-----AANLSLESRSQALTSV 129
Oy 142 -----OGKLVESAATATSSSTSNSHI---HQHSDSNSSALPHH1SSSSSSNNSSG 192
Db 130 VSETRARQAVTTANASATISTVTATETATATASTAATSEALIDD-PSAINTNNNN 188
Oy 193 NRAHIAAASARATPAATPANSELYKLTORAKMTSMDMAQLAOFSLADFNLIN 252
Db 189 SKAQADASESYK-----TKVISYHSEDOOQOQOQOQOQOQOQOQOQOQOQOQOQ 242
Oy 253 SLASQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 298
Db 243 HQQAQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 302
Oy 299 DL-----SSKSPNSIS-----GDV-----KSVRCATPTPSGRAYSEE 334
Db 303 QLYEYETDTPMRNFAFYVTGGGGVGMVMTSHEHLAATAAATAAAG-----ITP 358
Oy 335 DLSEALQDVANKLDARKSASQHEORSILD-----NRLFKMKHDOEDDH 380
Db 359 NIDEVIQDTLKDCEDEHSTYVHLTVSDMHLTKDSSPYALTNEQLHQOHHQOQLH 418
Oy 381 DQDELEDSDNDAEAEVDSNASTPVYPAEPARQLKLSLSEHNSDGLGEDVDRCSPKMG 440
Db 419 HHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 453
Oy 441 RHPAGNA-----SANOGAPASIPLDANLHTLMLAGISMPKLDDETQYGD 489
Db 454 DSPSSHLSTLQSTQTLTSATQKDSLPEND-----YRAAQLGS-----SLQNSV 502
Oy 490 FIKGLLVANSGIM-----NEGLNLTLASQENSGNASLLQOQOQHQQHQQHQQOQO 543
Db 503 YAGSLLTGTANGIQYGMQSPNQTQHLQOQHQQO-----QOQOQHQQOQOQOQOQO 553
Oy 544 QOQHYAARHRLPKSETPETNSSLDPNDASEDPILKITSFKVSGFA-----SSSLS 599
Db 554 QOHHNQOQH-----NSSSSP-----GPAGLHSSSSSAATAAA 588
Oy 600 VGGHHPLNNNSLSISNNSHSSNHRNGNRS-----PHSAPMLAAVAAG 649
Db 589 VAAATAAVNGHNS-SLEDYGSFRSSHSGGGGCTLPAPORAIYNSGSEVERIATIT--N 645
Oy 650 YSAGN-----SLTSSSSSIQ-----KMAANIOROINQSGQESLRNGNVSDCSN 696
Db 646 YRGQNDTWFDPLSYATSSSGOQLGVGAGVSNV-----IRNGRAISANA 693
Oy 657 NGSSSLG-----YKPSISYAKII-----GDTDSRFPASPNLSQOHSANH 740
Db 694 AAAAAADGTGRVDPGTFLSASASISATLFPADYFTESECEVCNCAISTPLRRDNTGHY 753
Oy 741 LTHQ-----QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 785
Db 754 LCNAGCLYKANKGNRPLIKOPRRISASRAGLSCSNCHTTTSTLMRRN-----DAG 805
Oy 786 EMSVHRAGSYGVSTLEYKVEKHEHLMRPRKEBEPQPDVLGLTGPANKQLDYLKAGP 845
Db 806 EPCVACGLYYKL-HSVPRPLTMKQDTIQKRKRK-----CTSEKSKSKSKNAL 855
Oy 846 HGGSLNALXKNNQAAAAAATAATNGCLKPLFEAGP-QALSFOPNMFWD 904
Db 856 NAIMSGSLVTNCHN-----VGVLDSQMDVNDKMPQLDKFYNSYSQPOQLPQ- 908
Oy 905 NATNAYGLDFNRITAMRNPOASNHGLKMSAQDVENVYDGIIRKTIQASGNSAAGN 964
Db 909 -----YQOQOQOQLMADQSSAASSPSM----- 931
Oy 965 GSNGSNGNGHGHG-HGHALLDQL-----VKTPPLFTNHRNDYATC--- 1009
  
```

Query Match 5.0%; Score 302; DB 5; Length 1140;
 Best Local Similarity 18.4%; Pred. No. 8.8e-09;
 Matches 250; Conservative 175; Mismatches 411; Indels 522; Gaps 56;

```

Db 932 GSTLSPSAMSHQHTPHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 988
Oy 1010 -SSASGESYKRSRGSPMGNYADIKERKRLSADSGGSDDEHSAHSHNNNNNDLAHKXKSG 1068
Db 989 HSPSTPSIFNTFSP-----THQLHNNNN-NNNNSSIFNNNNNS 1028
Oy 1069 GGGGGGNGQTNNGNRRSRRMSTRDS-----ETD 1096
Db 1029 SSNNNNKLIQYKIQAOQOQSSSSSGSTSDHQLAQLLPNSITAAAAAATAAIAKTE 1088
Oy 1097 ASSFKSGENGQOQNHKMDLNGSSSSSHIKCSEA-----ATGHH-----SP 1139
Db 1089 ALSTSQANCSATAGLAVTSTPTTASSTLSLSHNSIISLPNHOAGMTLCKPRTSP 1148
Oy 1140 GHHTTSLHEKLAQIKAEVDQADQ 1164
Db 1149 PYLTPBEDQPALIKMEEMDQSQ 1173

RESULT 15
O8T1T0 PRELIMINARY; PRT; 1140 AA.
AC O8T1T0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical 124.5 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafarski K., Pachebat J., Dear P.,
  Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
  Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
  "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
  RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AC116102; AAI93607.1;
DR InterPro; IPR00198; RhOGAP.
DR Pfam; PF00620; RhOGAP.1.
DR SMART; SM00324; RhOGAP.1.
DR Hypothetical protein.
KW Hypothetical protein.
SQ
SEQUENCE 1140 AA; 124451 MW; 785FAF46D83DC4CC CRC64;
  
```


QY 791 RAGSYGVPHSTLEYKVERHLMRPKREPKQPDVLGTLGPANKLQDLKXAG--PHG 848
DB 904 LAGSTSGSSSTFYGSSDSR-----DGSVSTGSSSNTDASTDLAGSSTSG 950
QY 849 SKLSNALKNQNNQAAAAAATAATPNGIKLPLFAGPQALSFQPMFQPTNATN 908
DB 951 SSTYGVSSSRDGSVSTSSSNTDASTDLTG-----SSTS 986
QY 909 AVGLDENRITLQAMRNQASNNHGLMKASQDMWENVYDGIIRKTLQASGNGSAGNGSN- 967
DB 987 G-----GSSTYGVSSDSRD-----GVSSTGSSSNT 1012
QY 968 -----GNGNGHGHGHGHGHALLDQLLVKTPLPFTNRRNDYAATCSA-----SG 1014
DB 1013 DASTDLAGSSTSGSSSTFYGSS-----SSNRDGSVATGSSSNTDASTTE 1056
QY 1015 ESVKSGSPMGVADIKRE-KLSADSGSSDEHSHASHINNNSDLA-----NKXK 1065
DB 1057 ESTTSGSSTEGSSSSHDGVSITDGSSTSGASSASATKSDAASEDEGFMMWNRK 1116
QY 1066 SGGGGGGGNGQTNNGRFRSRRMTSRDSETDASSFKSGNGGQGNHKKMDLNGSSSSSH 1125
DB 1117 S-----GSHKATVQSSITDKTSDASSTDTSTSGASTTSGSS--STSGSSTSDA 1170
QY 1126 IKCESEAAATGHSRPGHHTTSLHE 1149
DB 1171 SSTSSSVSRSHS---GVNRLHK 1191

RESULT 14
F90073
Hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #ext_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hatori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.;
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KIR>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447

Query Match 4.5%; Score 269; DB 2; Length 2271;
Best Local Similarity 17.2%; Pred. No. 1.5e-05;
Matches 195; Conservative 200; Mismatches 506; Indels 236; Gaps 27;
QY 43 IAIAGS---EDPPOYNHSSKEISQSNPNHCKTENHRLQCHNGSOLLEEDSENNTQSH 99
DB 1015 ISTSGSLASDSKSMVSVMSTSGST-----SESLSDQSTSDSDSKLSLS- 1064
QY 100 DSSRTPTGATSTPPEPPIDWRPSAKCNFCVNGRLTVNAQGLVAESAATATSSSTS 159
DB 1065 ---TSOGSSTSTST-----STASVRTSSQSTSGSMSASQSSMSISTSFSDS 1110
QY 160 NSHIHQHSDSSASLPHHTSSSSSSNNNSGNARHIAAASATAATPAATPANSLEY 219
DB 1111 TS-----DSKASASTAS-SESIISQASASTSGSVSTSTSLSTNSERTSTVSVDSTSLSTS 1164
QY 220 KLTLORAAKMTSMDSMAAQLAQSILADENLINSLASQOQQOQQOQIASAVPTPSEVA 279
DB 1165 ESDSISESTSTSDSISEAISSESTISISBSNSITSDSESQASAFSLSGSESTSEST 1224
QY 280 AAI-----SPALKDTPSPVDAPLDLSKSPNSISQDVKSVACATPTPSGRRAVSE 333

DB 1225 ESVSSSTSESTSLSDSTSESGSTSTSLNSITSGASIS-----TSTSIESTSTFKS 1276
QY 334 EDLSRALQDVVANKLDARKSAS---QHHEQNSILNRLFKKHHDQEDHDQDELED--- 387
DB 1277 ESVSTSLMSNSTSTSLNSNSTSLSTSLSDSTSDSKSDSLSTNSSTSDSTSTSKSDSTST 1336
QY 388 -SNDAAEAEVDSNASTPYPAEFARQRLKLSLSEHNGSDLGEDVDGSPKMGHNPACG 446
DB 1337 LSGSTSESESDTSS-----SESKSDSTSMISMSQSTSGSTSTST 1377
QY 447 NASANQAPASLPDLANVLHTMLAAGIAMPKLDFTQVGDPIKGLLVANGGIMNEG 506
DB 1378 STSLSDSTSTSLSLAS-----MMS 1398
QY 507 LNLTLASQENNGNASILLQOQHQQHQQHQQOQQOQVAAVRRH-LPKSETPENS 565
DB 1399 GVDNSASQASNSTST-----STSESQSTSTSTSTSTSTSTSTSTSTST 1443
QY 566 SLDPNDASEDPLKIPSFVSGPAPASSSLSPGLVGHNNHPLNNNSLSISNNSHSSNS 625
DB 1444 SLSDS-----TSISKSTSGSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1493
QY 626 HRNGNRNPHASAPMLAAVAAGYSAGNSLITSSSSSIQKMMASNIQRIQNEQSGSL 685
DB 1494 SLSDSTSTSGSASTSTSLSN--SAGASESDSSTSLSDSTSTSTSTSTSTSTST 1550
QY 686 RGNVSDGSSNNCGSSSLGYKKPSISVAKIIIGDTPSRFGASPNLISQGHSHAHLLTHQ 745
DB 1551 LSDSLSTSTNR-----MTIASLSTVSTSGSSTSESTSDSTSTSLSDSQ 1599
QY 746 Q-QOQLSAQEAALGKTRPKRGRKRYNRVDRSLVEAVKAVQRGEMSVHRAQSYGVPHSTLE 804
DB 1600 STSRSTASGASAST-----STSDRSTASSTSTSMR 1632
QY 805 YKVERHLMRPKREPKQPDVLGTLGPANKLQDLKAGHGGSKLSNALKNQNNQAAA 864
DB 1633 TSTSDSQSMSTSTSTSTSTSTSTSTSDSTSTSTSTSTSTSTSTSTSTSTSTST 1692
QY 865 AAAAAAATAATPNGIKLPLFAGPQALSFQPMFQPTNATNAYGLDFRITAMNPNP 924
DB 1693 ASEVMSASISQSNSESVNDESESVSESNESDSK--SMSGSTSVSDGSLSVSTSLKS 1750
QY 925 QASNNHGLMKASQDMWENV-----YDGIIRKTLQASGNGSAGNGSNGNGHG 975
DB 1751 ESVSESSSLSSQSMDSVSTSDSSLSVSTSLASSEVSJSDSLSDSKSTSGS----- 1804
QY 976 HGHGHGHALLDQLLVKTPLPFTNRRNDYAATCSASG-ESVKSQSPMGVADIKRE 1034
DB 1805 -----TSTSGSLSTSTSLSGSEVSESTSLSDSIS----- 1836
QY 1035 LSADSGSSDEHSHASHIN-----NNNSDLANKKSGGGGGGNGQTNNGRGRSRM 1087
DB 1837 -MSDSTSTSDSDSLSGSTSLSTSLSDSLSDSKSLSSQSMGSG-STSTSVSDSQS 1894
QY 1088 TSDRSETDASSFKSGNGGQGNHKKMDLNGSSSSSHIKESBAATGHSRPHCHTT 1144
DB 1895 STNSQPTDSMSIASBESDSMTSDSNISGSNSTSTSLSTSTSDSMGSSVSVSTST 1950

RESULT 15
S66149
gene p1psequae protein A long form - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #ext_change 23-Sep-2002
C:Accession: S66149; S66150; T45461
R:Weber, U.; Siegel, V.; Mlodzik, M.
EMBO J. 14, 6247-6257, 1995
A:Title: p1psequae encodes a novel nuclear protein required downstream of seven-up for t
A:Reference number: S66148; MUID:96134923; PMID:8557044
A:Accession: S66149
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1085 <WEB>

[illegible]

RESULT 13
S52714
sericinIB - silkworm
C:Species: Bombyx mori (silkworm)
C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
C:Accession: S52714

R.Garel, A.A.; release, G.G.; Prudhomme, J.J.
submitted to the EMBL Data Library, March 1995
A.Description: Structure and organisation of the Bombyx mori sericin I gene and of the flanking regions
A.Accession number: S52714
A.Sequence: 552714
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-1217 <GAR>
A.Cross-References: EMBL:Z48802, NID:g7556699, PID:g755700

Query Match	4.5%;	Score 270.5;	DB 2;	Length 1217;
Best Local Similarity	18.5%;	Pred. No. 5.7e-06;		
Matches 238;	Conservative 174;	Mismatches 483;	Indels 389;	Gaps 45;

0y	5	SYEISLEVEAECECMGRPMKHYOKPLTCSHLINEEOQI	I	A	L	A	G	E	D	E	P	S	Q	-----	YN	56										
			:	:	:	:	:	:	:	:	:	:	:													
			:	:	:	:	:	:	:	:	:	:	:													
Db	158	SYGVASADSSGASSRRRQANYSDK	-----	D	I	T	A	S	K	D	D	S	A	D	S	205										
			:	:	:	:	:	:	:	:	:	:	:													
			:	:	:	:	:	:	:	:	:	:	:													
0y	57	H-----SSKEISOSNPNHCKTEN--	H	R	E	O	O	H	N	G	-----	S	O	L	E	E	D	E	S	E	N	O	T	-----	97	
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	206	RDSDESEAGSLDRSSASSKNDNFVYRTKOS	I	G	O	A	K	S	R	S	S	H	O	E	B	A	D	A	Y	N	S	P	D	S	Y	265
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
0y	98	---SHDSRTTPCATSTPSPPEPIDRPAKCNFCVNGRL	L	T	N	A	G	K	I	V	A	E	S	A	A	T	A	T	-----	154						
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	266	NAGTDSSTSKKKASSTTVADKOI	-----	B	A	N	D	R	S	S	K	O	L	O	S	A	I	S	-----	311						
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
0y	155	S-----SSTSNSHIHODSDNSA-	-----	S	L	P	H	I	S	S	S	S	S	S	N	N	S	G	N	R	A	H	I	198		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	312	SGPKGTSVSXNDROYSNDRKSDAIVYGRDGTVA	Y	S	N	M	D	S	E	K	T	S	I	Q	S	T	N	T	A	D	O	N	S	V	R	371
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
0y	199	AAASARATP-----AATPANSLELYKLTORAAK	M	T	S	M	D	S	M	A	A	O	L	A	O	F	I	L	A	-----	246					
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	372	SAASDOTSKSYDRGYSDNUIVAHSGSGSONKSS	Y	R	A	K	D	G	S	S	T	N	E	K	S	F	S	-----	428							
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
0y	247	DNILINSASQQQQQQQQQIASAVTPTTSEVSA	A	I	S	P	A	L	D	T	P	S	P	V	A	D	L	S	S	K	P	S	306			
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	429	-----SSNSVETSDGASARESA-EDT	-----	K	S	S	-----	454																		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
0y	307	NSISGDVKSVPACATPTPSGRAYSEEDL---	S	R	A	L	O	D	-----	V	A	N	K	L	D	A										

Db 292 SSTSTVNTCCAANNNSNNNNNNNNNNSTGSLCALIMDRSPDGNINTITEESTGNNSPRS 351

Qy 793 -----GSYGV-----HSTLEYKXERHLMRPRKE---PKDPDLVGL-TGPA 833

Db 352 NCGSNCGSGNGIPLSPRLSLSSNGSVNNSPRNIHLNMLNNNSNPLSPRIINHIWVS 411

Qy 834 NKLQDLKIKAGPHGSKLSNALKNONNAAAA-----AAAAAATAATPGLKLPLEFA 888

Db 412 NLNNNNNNNINPNNNPNNSN--NENNNVSPRNNHNISPRCSNISPRSNNGSTTI--- 465

Qy 889 GPQALSPQNNFWPOTNATNAVGLDFNRKITEAMRPQASNNHGLKMSQODVENVYDCII 948

Db 466 SPRNISNNNNI-----NNNNNNILTPPHNSPRLFNVPNTSPRLATISLNTLPVSSI- 521

Qy 949 RKTLOASEGNSAAGNGSGNGCHGHGHGHALLOLLVKYKPLPFTNRRNDVYAT 1008

Db 522 -----TSSNNNNNSNNNTNPTSLNNNGR-NGHCIOITISEILGNKPVVYNNNGNNNNNT 575

Qy 1009 CSSASGESVKSRSBPGWYADIKREKLS----- 1038

Db 576 NNSSTTNNNNITTTNNNNNNNNNNINNVLTSPRKRTGKNHSTKTSLDQFETSSNGGDSIS 635

Qy 1037 -ADSGGS-----SDEHSASHIINNNSDLAHNKVKSGGGGGGGNGOTNGRSSRMTS 1089

Db 636 GAGSGGSLRRRNKODNDENDGNSNTNNSNNNNNNNNNSNNNNNNNNNNNNNNNNNNNN 695

Qy 1090 RDSFETDASFSKSGENGQOHNKMDLDCGSSSSSHIKCESAALGHHSPG 1140

Db 696 NNN 744

RESULT 10

forked protein 5.4K - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 22-Jun-1999
C:Accession: S57238
R:Hoover, K.K.; Chien, A.J.; Corcoran, V.G.
G:Genetics 135, 507-526, 1993
A:Title: Effects of transposable elements on the expression of the forked gene of *Drosophila melanogaster*
A:Reference number: S57236; PMID:94063487; PMID:8244011
A:Accession: S57238
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1436 <HOO>
A:Cross-references: EMBL:X69871
C:Genetics:
A:Gene: FlyBase:f
A:Cross-references: FlyBase:FBgn0000630
A:Protons: 43/2, 104/3, 164/3, 564/3, 717/3, 825/1, 1001/2, 1121/3, 1204/3, 1274/2
A:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
A:0-83/Domain: ankyrin repeat homology <AN1>

	Query Match	4.6%	Score 275;	DB 2;	Length 1436;
	Best Local Similarity	19.2%;	Pred. No. 4,2e-06;		
	Matches	210;	Conservative 146;	Mismatches 369;	Indels 368; Gaps 39
OY	22 QMKHYQDKLTCTSH-----LWIEEOPIAIGSEDEPSOYNHSSKEISQ	64			
Dd	136 RWRGSRKSLDVKYGKSPINDAEANOVECLAVLVQHGTSTVDYNGKSSQRHKSDQDLHQ	195			
OY	65 SNPNHCKTENRRLLEOHNGSOLLLEEEDSNNTTSHDSSTPT--FGATTPPPPEPIDM	122			
Dd	196 -----QHQQQQQQQQQLHLSSCSNNSNSSKOTSSNTTKSKSSSTLLSDVEPYL	246			
OY	123 RPSAKCNFCVNRLLTVNAOGKLVAESAATATTSSTSNHHIQHDSDSNSASLPHTHS	182			
Dd	247 HPPALAGSGCGSMGLGMCMGMCMGMSWLGISTRKNSDA---LYSQGSNSESXKLYNGSS	303			
OY	183 SSSNNNNSGNFARRIIAAASAATPAATPANSLELYKLLTPRAAKMTSMDSMAOLAOF	242			
Dd	304 SLGGNSSSGCG-----GVGGGGGSGGALLPNDGLVNPV---RNGGMVTTPPNGISIGE	355			

```

QY 243 SL-----ADFULINSLASOQOOOQOOOLASAVTPPTTSEVSAALISPAKDOTSPSYDAPL 298
Db 356 SFFLDHPDYHLQSGAGSVGRGLRLOOOCETAQO----- 388
QY 299 DLSKPENSSISGDKSVKACATPPTPGRAVYEEDLSRALODVVANKLIDARKASOHN 358
Db 389 --ESFPRAAYTESGORVTHQADVTAPA-----DSGSDESVSISASTRRSSSSC 438
QY 359 EORSILDRFLFKMHQHDOEQDHGDLEDSNDDAEVDNSASTPVYAEPFAFOQLKUS 418
Db 439 AARVASI-----ATVAATAASP-----G 455
QY 419 HISENGSDLEDVDRCSPKMGRHACGNASANOGAPASIPIDANVLHTMLAIGAM 478
Db 456 NIKNNN-----YKTFRASP-----AIRMAAI-----477
QY 479 PKDETQTVGBFIKGLLVANSGGINEGLNLILSASOENSNGNASLLLOOQHOQH--OQ 537
Db 478 ----TTT-----EDIYLRESRKHQOQOOQOHOLOOQOOLOHQATO 514
QY 538 HHOOOQOOOHAAHYHRLPKSFETPTETNSLDPNDASEDTLKIPSKVCGSPASSSSLG 597
Db 515 QOOOQOOOQOOOLOHKYXWGRSKRSRGSR-----SRASASSTRT 555
QY 598 GLV---GGHHPLNNNSSLSISNNNSHNSSHRNRSRHSGASPMU-----641
Db 556 DIVLOYSNH--LNKRMMNNNSNNNSNSSNIAGSSSSNNNNNSLNRNKSHSIGLSH 613
QY 642 ---AAAVALGGSYAGNSLTS-----SSSIQKMASTIQOINQSOESLRNGVSD 692
Db 614 SKYESCLDNYSAXKNVNLKNOLLNGIKSIDYESVCPRPDVAERTQOTHKSNIRNNLD 673
QY 693 CSNNNGSSSLCYKKPSIS-VAKITIGDITSRFGASPNLISOCHSAHNLTHOQOOOLS 751
Db 674 ASSNNNTGSGIN--NISNIGNMNGNOSSR-----NLKRVSASRPOMNL- 715
QY 752 AOELAGKSTRPKRGKYRYNRDSDLVEAVKA---VORGENSVHRAGSYGVGPBSTLEYKV 808
Db 716 --AVVNPPR-----PRLPPRLARPVOQRNNLSVOPNS-----748
QY 809 ERHLMRPKREPKRPORDLVGLTGPNANKLODLTKKGPHGSGKLSNALKNONNOAAA-- 865
Db 749 -----MTGPT--MHLQKVGCPNGP-----TNANGCGAAPR 780
QY 866 --AAAAAAAAAATNGUKTLPLFEAGROALSFOPMFMFOTATYNAYGLDFNRITEAMRN 923
Db 781 VPAPRPVATEAVDSOSGLEV-----VEBPSLRPELVBGNHRTWTSTISANKKALKL-N 833
QY 924 PQAASHHGKLMSAQMVENYVDGITRIKTLQABEGNSAA-----GNGSGNGNGH 974
Db 834 AGSTNGSSIIAASDDS-QSRYGGSVH-----AANGSANGHFYGYSEGNKONGXNANGN 886
QY 975 GHGHGHGHALLDQLLVKKTPLRPFTHNRNNDVATCASAGESIVKRSGBPWNADIKRER 1030
Db 887 GNGNGGGG-----NYHSHPNQPHVAT-----GQPHQYPSLYG-----919
QY 1035 LSADSGSSDEEH 1047
Db 920 ----GSSAEHDH 927

```

RESULT 11

forked protein 5.6k - fruit fly (*Drosophila melanogaster*)
C.Species: *Drosophila melanogaster*
C.Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 22-Jun-1999
C.Accession: S57237
C.Hoover, K.K.; Chien, A.J.; Corces, V.G.
Genetics 135, 507-526, 1993
A.Title: Effects of transposable elements on the expression of the forked gene of *Drosophila*
A.Reference number: S57236; WUID:94063487; PMID:8244011
A.Accession: S57237


```

Db 654 QQLQQQQQQQQQQQQQQQLHQQQQQHQHQHQHAIQHQQSQQASKISYR 713
Qy 785 GEMVHRAAGSYGVPHSTLEKVKERHLMRPKREKQPOLVGLTG---PANKL--- 836
Db 714 GIFTT--TGNMNNMAAAAAAQQHQQQQQHQQQQLPSPOLGLVAGSPSPNSLGSW 771
Qy 837 ---OLDKLKAGP-----HGSKLKSNLKNQNNQAAAA-----AAAA 871
Db 772 GLRSPDKTFQPPPLFSLPAHYATMQQQQQQQQQQQQQQAAAGALSPFDDGAAAA 831
Qy 872 AAAAATPGLKTL---PLFEAGP-----QALSFQPMF 900
Db 832 AAQAHELGLTMDCTPLLLKQPPPYAGASAGPGLGDLHSSHEQQLQQQVYRSQPKYQ 891
Qy 901 WPGNATNAYGLDFRITRTEARNPQ-----ASNHGMLKSAQDMENVYDGI 947
Db 892 WLDSPAD--YAQQQQQQVQVQVQVQVQVQVLPGLPTSSASSSMAALGVLPKQENYD- 947
Qy 948 IRKTLQASEGNSAANGNSGNGNG-----HGHHGHALLDQLLVKK 992
Db 948 ---MQPS-SNGTGYSGSGSSGSSAAAAAANAVALAETSPSTKGEILSGVYQS 1002
Qy 993 T-PL 995
Db 1003 TVPL 1006

```

RESULT 7

```

113349 parallel sister chromatids protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13349
R:Gandhi, R.L.; Goldberg, M.L.
submitted to the EMBL Data Library, November 1995
A:Reference number: 217657
A:Accession: T13349
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1768 <GAN>
A:Cross-references: EMBL:U40214; NID:g1100982; PID:g1100983; PIDN:AAA91230.1
C:Genetics:
A:Gene: pasc
A:Cross-references: FlyBase:FBgn0004655

```

```

Query Match 4.7%; Score 282.5; DB 2; Length 1768;
Best Local Similarity 19.4%; Pred. No. 2.3e-06;
Matches 238; Conservative 160; Mismatches 467; Indels 361; Gaps 45;

```

```

Qy 48 SEDPSQVNHSSKELISQSNPNHCKTENRLEQNHGSQL--EEDSENQNTSHDSSRPT 106
Db 203 ASEQEPQSSASGKQKQKPK---EEKKLKPEAPPSVLGRARAVYREVDEDERYPT 258
Qy 107 PGATSTPPEPEPIDMRPSAKCNFCVNGRLLTVNAQKLVASATAT--SSSTNSHIHQ 165
Db 259 P-----TKDLIIPKAGQPAEVAATATLAAASSAFSS 292
Qy 166 HDSDNSASLPHHISSSSSNNSSGNRAHIIAASAPATPAATPANSLELYLTQR 225
Db 293 TFGSGSEPSLPPPTSPASASASTSSQLPS--ASGSASNPSPASRTPREPHIVLRISKG 349
Qy 226 AAKMTSMQMAQLAQLFSLDNLINSLASQQQQQQQQLASATPTTSEVSAALISPA 285
Db 350 TSLRVSTVS---EEPPSSPAHQNOLNOLSTVEEPARSGETVPASTPKTT---VKPL 403
Qy 286 LKDTSPSPVDAPLDLSKSPSPNSISGVKSVACATPTPSGRAYSEEDLSRALQ--- 341
Db 404 RPTTADVD-----GSSAAVGCASAGDSFEERKQSLPEPDEDEDEDEDEPEPEI 458
Qy 342 -----DYANKLARKASQHEQRLNDNLFPMKHHDQ 378
Db 459 NYCTVKGSPDKPKERLKLIIKTVDIYRNAIAKAAAA---ESRSSEKSKRSKHGH--KQL 513

```

```

Qy 379 DHDQDELDSDNDAAEVDNSASTP-----VYPAFAARQKLKSLSEHNGSDL 428
Db 514 LAAGSAAAPSGATPAEINSEFTSPHALLSEANSQQQOHTSHLQHLNRPQBSAV 573
Qy 429 -----GEDVDRGSPKGRHPACGNASANOGAPA----- 456
Db 574 ISPTTRDHDHFDSSQSVLGSISSEKGNSTPOLLAQVQEDSCVIRSGSVITSDLETSQH 613
Qy 457 -----STPLDANVLLHTLML--AAGICAMPKLDQVGVPIGLLVANGGIMNEELN 509
Db 634 SSLVAPPDSDIESRLSEMMMTIDAGTGAASAVETP-----LQEDILA 676
Qy 510 LLSAQNSNGNSALLLQQQQHQHQHQHQHQVAAVYRRLPKSEPTNSLDP 569
Db 677 VLREVEVRLGNTD-----PEPTEEDQQQPKKATRGKKA-----NNVDV 720
Qy 570 NDASEDPLKI PSFKVSGPASSSLSPGLVGHHPHLNNNSLSISNNSHSSNRHG 629
Db 721 TPATE--TRTGRKAGADATTAISP-----PTGKRN-----TRGTGRSKRAQEV 765
Qy 630 SNRSPHASPMIAAAVQ-----GYSAGNSLITSSSSSIQKMMASNIQRI 676
Db 766 DMEVDETAMTTPVPAHEQLEQATLPPRRGRNAAARANNNNININIAAN-----L 820
Qy 677 NEOSGESLRNGVSDSSNNGSSSLGYKKPSISVAKIIGTDTSRFGASPNLLSQOHH 736
Db 821 SAKAEASRLAEGYA-----GGAARSYGRKAKQOQTVQL-----QOE- 858
Qy 737 SAHLTHQOQQQOLSAQALGKGTTPRGKYRNYDRDSLVEAVKAVORGEMSVHRAGSY 796
Db 859 -----PVBEQETDADE-----EQTPRAK----- 878
Qy 797 GVPHSTLEKVKERHLMRPKREKQPOLVGLTGPAKTLQDLKAKP-----HGSKL 851
Db 879 -IPHTD-----HRBSPDHPDPPDELSSNNSSSLQHDGSSSSPPRDFKFKDKF 929
Qy 852 SNLKNQNNQAAAAAATAATPGLK--LPFEAGPOLSPQPMFWPQTATN 908
Db 930 KRTITLDTQGAANAGAGAAAAAPRESSGEORGAVKLVISKKKSIKSRALVSDQAEQ 989
Qy 909 AYGLDFRITRTEAMNPQASNHGMLKSAQDMENVYDGIIRKTLQASEGNSAANGSNG 968
Db 990 A-----TVAKKH-----LYKHSW-----ALEANG--CGTNSDA 1017
Qy 969 SNGNGHGHGCHGALLDQLLVKTPLPF---TNRNDVIAATSSASGESVKSQSP-M 1024
Db 1018 SNASAGVYAGAKDHLHLAAGKSDGDFGSPSSNNGSSSACSSA---STLRGDSPAL 1074
Qy 1025 GNYADIKRERLSADG-----GSDDEHSASHINNNSDLAHNKNK 1065
Db 1075 G-----KISRLAGQGVPAITSTSSDADLDLEPIAGELDLERSAA-----GA 1116
Qy 1066 SGGGGGG--GGNGQTNGG--RSSRMTSRDSETDASSFKSGENGQGNHMKMDLNGSS 1121
Db 1117 SAGGTGATTGGGATGGGPRVDRKTK--DYEVVANVKA-----HQIOIGEXYOE 1167
Qy 1122 SSSHICESEATGHSFGHHTTSL 1147
Db 1168 MDDDEYILDALQPHNPATRCUSAL 1193

```

RESULT 8

```

113610 parallel sister chromatids protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13610
R:Murphy, L.; Harris, D.; Bartell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217668
A:Accession: T13610
A:Status: preliminary; translated from GB/EMBL/DBJ

```

```

Db 97 000000HRNSISNMDRTVNAKFKPASSVAGGPNNEVRNNSMLTPSRGVITGCTGNIRK 156
Qy 218 LVKLLTORAKMTSDMSAAQLAOFSLADFNILNSLAQ000000-----QQIASAV 270
Db 157 LRV-----SLLSNHFAVCYPRPSNIYONSNAGNSALQRTTSSLRLNMSRVAADA 211
Qy 271 TTTTBEVSAAIISPALKDTPSPSVDAPLDLSK-----PSPNSISGVKS 316
Db 212 TPTT--VSRASSNSISLATSTSL-APKSSSSSGSGNSTPQ0000QVLSSNNSSSSNS 268
Qy 317 VAACATPTPSGRAYSEEDLSRALQDVANKLDARKASQOHQESILDNRLFKNHQ 376
Db 269 FTKASSPNNGARSVCGAATSAATGTTA-----AGSHHQPHHHHHHHHHQHNNH 321
Qy 377 EQ----- 378
Db 322 000000TSLSGHSLITVAGSASAGGGGGSSSGTAAGCTNRKPKTTSSPFTSVT 381
Qy 379 -----DHDGELEDSDNDAAEV-DSNASTP-----VYPAEFARAO 413
Db 382 VGHPRKLNAGDTGDESADLDESHDDNSRITDLENETPSMEDPFSKEEYVYANNAALST 441
Qy 414 LRLKSLHSNNG-----SDLED 431
Db 442 NARVLPSTSQYGLVVVDPIAPSLGOTIQNVQVNSDNIIVVSGAVTPGCTKKDKDET 501
Qy 432 VDRGSP-----KMR-----HPACGNASANOAPASIPLDANVLLHTLML 471
Db 502 QHRSRFPKVKTESTEPFKGRMCMCDYLDHSSVGNCGNNKETSSTSEALAA-TDGG 560
Qy 472 AAGTGA--MKLDETQTVGDFIKGLVANSQGINNEGILLNLSAQSNSNANSLILQO 528
Db 561 AAGVAGSEAPRAHKTQS-----MLPPTQKINE--NHEANSTANNVNAVEQQO 609
Qy 529 00HQH00H-----H0000000HVA 549
Db 610 00000000TIVGNAITKTLPVALRNVSRSSSVTRSPNATVEFLPNLLA00000000L- 668
Qy 550 AYRHLRPKSETPETNSSLDPNDAE-----DPLIKIPSFVK 585
Db 669 -----FDSVNAASAASPFPACDPNNNDYARTAMOLHOTLOOKOREDMAYP---- 716
Qy 566 SGPASSSSSLPGG--LVGHHNPLNNNSLSISNNSNHSNSH--RNGSNRSPHASPWL 641
Db 717 PGAGGYANYQNGSDAVGA-----ASNNSAAATGESQLSYVGG000000PLSPAPLT 772
Qy 642 A-----AAVAGGSAGNSILTSSSSSIQKMAASNIQ-----RQINEGSGGSLKNGVNS 691
Db 773 PQADPTFAAVAG--QSPNFQLE000000QATSQIDGIVPQPPNPO000000TPOQSTA 830
Qy 692 DCSNNGSSSLGYKPKSISVAKIIGTPTSRFGASPNLLSQCHSAHHLTHQ000000L- 750
Db 831 QQAATAAATSAVTAAPRPOGTSTNTSMAAVTTGOGOTMPLS-----HWTSE000PMLG 884
Qy 751 -SAQALGKTRPKRGKRYNDRDSLVEAVKAVORGEMSVRAGSYG-----VPHSTL 803
Db 885 AAAAAAAGGTAA-----TSVAAPQAI-P-TLQASAPSTIADPQQLWVPOQOQ 931
Qy 804 EYKVERHLMP-RKREKRPQDVLGVLGPRANKLOLDKLAGPHGSGSLNALKXQNNOA 862
Db 932 000H0E0000PQ0000PLPRANIASASANNNSNLMLTNTV-VATGEATNALTLDDQA 990
Qy 863 AAAAAA-----AAAATPGLKLPFEAGPQALSFQPNMFQPTNA--TNAVGLDFNRT 918
Db 991 TRLAALAAATGAAAAATGATSAAAATQOQIQLOQ00PNAESTESASGTSAAVAD-NKIE 1049
Qy 919 EAMRNPQASNHGLKMSAQDWE 941
Db 1050 QAM--DLVKSH--LMIIVAREVE 1068

```

RESULT 6
S69205

stripe a/b protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
 C:Accession: S69205
 R:Frommer, G.; Voithueggen, G.; Paeca, G.; Jackle, H.; Volk, T.
 EMBL J. 15, 1642-1649, 1996
 A:Title: Epidermal egf-like zinc finger protein of *Drosophila* participates in myotube gu
 A:Reference number: S69205; MUID:96203082; PMID:8612588
 A:Accession: S69205
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1180 <PRO>
 A:Cross-References: EMBL:U42403; NID:G1147788; PID:AA802355.1; PID:G1147789
 C:Keywords: alternative splicing

Query Match 4.7%; Score 283.5; DB 2; Length 1180;
 Best Local Similarity 20.6%; Pred. No. 1.2e-06;
 Matches 236; Conservative 128; Mismatches 457; Indels 323; Gaps 41;

```

Qy 35 LNIHQPIAIAAGSEDEPQYNHSSKEISQGNPNHCKTENHRLQOHNGSQLLEEDSEN 94
Db 3 LTRKQNEL-IVGSGQHSAATSSASASAGATSSPGLSQNLNVA--GAGATNTSTAN 58
Qy 95 NQTHSDSRTPPGATSTP---SPPEPIDWRPSAKNFCVNGRLITVNAQGLVAESA 151
Db 59 SODSLNTPPTTLGISRNPLOFAPPAPPIAVPSPAAGPFGYQYOTANAPRLHSPDA 118
Qy 152 TATSSSTNSHHQHDSDNSASLPHHSSSSSSNNNSGNNRARIHAAARATPAAT 211
Db 119 TSEVTAQPPV--ELDEYVDILQVQQLDSSA-----AAAA 154
Qy 212 PANSLLEYKLTORAKMTSDMSAAQLAOFSLADFNILNSLAQ00000000QIASAVT 271
Db 155 AANNPTTEGQVQPOQNTVQPHHQOQ-----QPOQ0000PQOVLAKPR 201
Qy 272 PTTSEVSAAIISPALKDTPSPSVDAPLDLSKSP-----NSS 309
Db 202 PRINLOKATEYAAQLAQVSSSPGSRVLLDYPSPYGVGNPHHTTPGEDLVALMFGSN 261
Qy 310 ISGDVKSARACATPPSGRRAYSEEDLSRALQDVANKLDA--RKASQNHQESILDN 366
Db 262 GTGGV-----APGTPSAM-----IMEGLTTLVAPTHNAFLLTETAAAHFNVLST 309
Qy 367 RLFPKAKHHQDQHDGDELSDNDAAEAVNSNASTPVYPAEFARAOJRLKSLHSENGS 426
Db 310 CLFK-----TSTAQSGSPRTATYISPAQFG-----SHGGA 343
Qy 427 DLGEDVDGSPKGRHPCGNASANOAPASIPLDANVLLH--TLMLAIGAMPKLDDET 484
Db 344 STSSN-----SLNSSSTAASSSTSSSLHYTTASAAAAAASAAAAASAN 386
Qy 485 QTVGDFIKGLVANSQGINNG-----LNLNLSAQENSGNSALILLOOQH00H0 536
Db 387 NSV-----LRAAQATATPTGSGSPGHVAVOPSAVASGRSSASHLSLNTSGHSPTSS 440
Qy 537 QH000000QVAAVYRHLRPKS-ETPETSNSLDPNDASEDDILKIPSPKVGSPASSSL 595
Db 441 AVEQEAHKQILEA-----LPGLDNTPTVTSDDISFPQPTTVVPRPTGSIEEDLSLE 496
Qy 596 P-----GGLVGHNHPNNS-----LSI-----SNNSNHSNHRGNSNPSHASPWL 641
Db 497 PQVIVASPVLSHSCSPLEERSTPPALAIYVESSNNSCNMYPOHNNNNNNNNNTTSSST 556
Qy 642 AAAVAGGSAGNSLTLSSSSSIQKMAASNIQROQINEGSGESLRNGVSPSCSSNNGSS 701
Db 557 TTTSKQ-----TTSENTCEVSPGNHTSQOQ--QOQLOHNTTSSNSGCHSH 605
Qy 702 SLGYKPKSISVAKIIGTPTSRFGASP-----NLDSQ-----HNSAHNLTH-- 743
Db 606 000000-----Q00QHMSFQ00YQ0QIILH0000PGYNHHHHHHHHHNSQL 653
Qy 744 -----Q00000LSAQEALGKTRPKRGKRYN-----DRDSLVEAVKAYOR 784

```



```

Db 361 RSLIDNLFKMKHHDDOEQDHDGDELEDSNDDAEABVDSNASTPVPYPAEFAPAOURLKLSHL 420
QY 421 SEHNSDLGEDVDGRGSPKMGRRHPACGNASANOQAPASIPLDVANYLHTLMLAIGAMPK 480
Db 421 SEHNSDLGEDVDGRGSPKMGRRHPACGNASANOQAPASIPLDVANYLHTLMLAIGAMPK 480
QY 481 LDEFQYDGFPKGLLVANSGGIMNEGLNLLSAQSQENSNGASLLLQQOQHQQHQQHQQ 540
Db 481 LDEFQYDGFPKGLLVANSGGIMNEGLNLLSAQSQENSNGASLLLQQOQHQQHQQHQQ 540
QY 541 QOQOQOQHVAAARRHLPKSETPETNSSLDPNASDPLIKTSPFVSGPASSSSSPGLV 600
Db 541 QOQOQOQHVAAARRHLPKSETPETNSSLDPNASDPLIKTSPFVSGPASSSSSPGLV 600
QY 601 GGHHPPLNNNSLSISNNSNHSNHRGNSRSPHASPMLAAVAAQGYAAGNSLTTSS 660
Db 601 GGHHPPLNNNSLSISNNSNHSNHRGNSRSPHASPMLAAVAAQGYAAGNSLTTSS 660
QY 661 SSSIQKXMASNIQROINEQSGESLRNGNVSDCSSNNGSSSLGYKKPSIVAKITGGTD 720
Db 661 SSSIQKXMASNIQROINEQSGESLRNGNVSDCSSNNGSSSLGYKKPSIVAKITGGTD 720
QY 721 TSREGASPNLLSQOHHSHNLTHQOQOQOQASOALGKGTGPKRGKXPNYRDBL----- 775
Db 721 TSREGASPNLLSQOHHSHNLTHQOQOQOQASOALGKGTGPKRGKXPNYRDBL----- 775
QY 776 -VEAVKAVQORGEVHRAQSYGV-----PHSTLEYVKEKSHMRPKREPRKPODLVG 828
Db 776 -VEAVKAVQORGEVHRAQSYGV-----PHSTLEYVKEKSHMRPKREPRKPODLVG 828
QY 781 .RCRELKCFRIERIVATMRTAFHTTGVQOGSTSDATIQ--ARQAAARSRRPRTQOAAAG 838
Db 781 .RCRELKCFRIERIVATMRTAFHTTGVQOGSTSDATIQ--ARQAAARSRRPRTQOAAAG 838
QY 829 LT--GPANKLQLDYLKAGPH-----GGSKLNSALKNNQNNQAAAAAAAAAATP 878
Db 829 LT--GPANKLQLDYLKAGPH-----GGSKLNSALKNNQNNQAAAAAAAAAATP 878
QY 839 QTEGGTTWMLQAEQCPQOPEPKQSGCGGCGSSSGRCHAQRPETAPFRGSGTVILSX-- 896
Db 839 QTEGGTTWMLQAEQCPQOPEPKQSGCGGCGSSSGRCHAQRPETAPFRGSGTVILSX-- 896
QY 879 NGLKLPLEAGAPOLSPQPNMFWPOTNATNAYGLDFNRITAMRNPOASNHG----- 931
Db 879 NGLKLPLEAGAPOLSPQPNMFWPOTNATNAYGLDFNRITAMRNPOASNHG----- 931
QY 897 -----BHVLADERHBCLR--PGLQSHHGDAEBPG 925
Db 897 -----BHVLADERHBCLR--PGLQSHHGDAEBPG 925
QY 932 -----LMKSAODVENVYDGIIRKTLQASBEGNSAAGKSGNSGNGHGHGHGHALLD 986
Db 932 -----LMKSAODVENVYDGIIRKTLQASBEGNSAAGKSGNSGNGHGHGHGHALLD 986
QY 926 LOSRLKKSADDMWGERLRMHQHQBADAGEQNGSLAAGNSGNSGNGHGHGHGHALLD 985
Db 926 LOSRLKKSADDMWGERLRMHQHQBADAGEQNGSLAAGNSGNSGNGHGHGHGHALLD 985
QY 987 QLVKTKTPLPTNHRNNDYAAATCSASGESVYKRSQSPMGVADIKRELSADSGSSDPE 1046
Db 987 QLVKTKTPLPTNHRNNDYAAATCSASGESVYKRSQSPMGVADIKRELSADSGSSDPE 1046
QY 986 QLVKTKTPLPTNHRNNDYVYTCSASESVYKRSQSPMGVADIKRELSADSGSSDPE 1045
Db 986 QLVKTKTPLPTNHRNNDYVYTCSASESVYKRSQSPMGVADIKRELSADSGSSDPE 1045
QY 1047 HSASHINNNSDLAHNKXKSGGGGGGGGNGGOTNGGRSSRMTSRDSETDASSFKSGENG 1106
Db 1047 HSASHINNNSDLAHNKXKSGGGGGGGGNGGOTNGGRSSRMTSRDSETDASSFKSGENG 1106
QY 1046 HSASHINNNSDLAHNKXKSGGGGGGGGNGGOTNGGRSSRMTSRDSETDASSFKSGENG 1104
Db 1046 HSASHINNNSDLAHNKXKSGGGGGGGGNGGOTNGGRSSRMTSRDSETDASSFKSGENG 1104
QY 1107 GQONHKXMDLNGSSSS 1123
Db 1107 GQONHKXMDLNGSSSS 1123
QY 1105 GQONHKXMDLNGGRADA 1121
Db 1105 GQONHKXMDLNGGRADA 1121

```

RESULT 2
A:33742
female sterile homeotic protein, 205k - fruit fly (*Drosophila melanogaster*)
N:Alternate names: membrane protein fish, 205k
N:Contains: female sterile homeotic protein, 110k
C:Species: *Drosophila melanogaster*
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 20-Sep-1999
R:Accession: A43742: B43742
R:Haynes, S.R.; Mozer, B.A.; Bhatia-Dey, N.; Dawid, I.B.
Dev. Biol. 134, 246-257, 1989
A:Title: The *Drosophila* fish locus, a maternal effect homeotic gene, encodes apparent men
A:Reference number: A43742; MUID:89276730; PMID:2567251
A:Accession: A43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2038 <HAV>
A:Cross-references: EMBL:M23221, NID:g157452, PIDD:AAA28540.1, PID:g157453

A:Accession: 343742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1106 <HA2>
A:Cross-references: EMBL:M23222
C:Genetics:
A:Gene: fsh
A:Cross-references: FlyBase:Fggn0004656
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: alternative splicing, transmembrane protein
F:1-1038/Product: female sterile homeotic protein, 205K #status predicted <MA2>
F:1-1106/Product: female sterile homeotic protein, 110K #status predicted <MA2>
F:59-116/Domains: bromodomain homology <BR01>
F:503-560/Domains: bromodomain homology <BR02>

```

Query Match      5.2%; Score 311.5; DB 2; Length 2038;
      Match Local Similarity 21.1%; Pred. No. 9.7e-08;
      Matches 236; Conservative 129; Mismatch 331; Indels 425; Gaps 50;

QY      140  NAQCKLVAESAATATSSSTSTNSHIHQHSDSNSASLPHHISSSSSNNSSSGRRARHIA 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1064  SASCKVEAYORANPVSSSSSSSS-----DSSSSSSSD-----SSSSDSDSDPAGDGE-- 1110

QY      200  AASARATPAAATPANSLELYKLTQRAAKTSMDSMAAQLAQSFLADFNLI--NSLSAQ 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1111  -----RPPKKKSRSDSNSNNVPSI-----NVVMGNNPS- 1141

QY      258  QQQQQQQQIASAVPTT-----SEVSAADISPAKDTPSPVDAPLIDSSKPS-P 306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1142  -----GALSPPTMLMGDLVHANSNTPISGMSMLGNANLPTAAALNNNNKTSLP 1191

QY      307  NSSISGVKSVACATPTPSGRAVSEEDLSRALQDY-VANKIDARKASQNHQRSILD 365
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1192  GSNFGG-----ADAPG-----NNMHAGAPVAGAAVSASTGQQHN----- 1227

QY      366  NRLFKMKHNDQHDHDEDESDNDDAEAVDSASTPVYPAEFA----- 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1228  -----KNGPNDISKQPPGPI-----NAALP--PHSFAGTATVATSQSSGI 1268

QY      411  --RAQRLKLSLSHNSDLGED-----VDRGPRKMGRRPACGNMANSAGAPASI-PL 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1269  RIASNHLKPSGL-----GGQDGEHNAALAAATGTSINSTG--TAGCGINNCGSNMNNANPL 1323

QY      461  ----DANVLLHTLMLAIGAMKEKDE--TOTVGEFTKGLLVANSGGINNIGLL-NLLS 512
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1324  GSGHGDMVNASLASLASLGKQKQIFPDPRVQSLAS-----LEPSASTGKSGLTDNPL- 1377

QY      513  ASQENSNGNSLLIQQQHQHQH--HQQHQQQQQQQVAAVYRHLRPKSETPEFTNSLDP 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1378  -MOOHLMPAPRQQQQQQQQPRGHQQQQQQQQQQQQQQQQQQQQQH----- 1421

QY      570  NDAEDERPLIKTSPKSVGGPASSSSLSPGLVGHNNHPLNNNSISINNSNH-----SNS 625
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1422  ----DYVTEL-----LSKGAENVGNMG--NHLNLFNLDMALAYQCKHPQQQQQA 1466

QY      626  HRNGSNRSPHSAPMLAAVAAGVSGANSLSLTSSSSSI-----QKMAASNIORQIN 677
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1467  HNNGFN-----VADFEWAFRGDGLNMTAAAFDLRLSPDLQQQMQQQMLQQQHN 1513

QY      678  EQSGQESLIRNGNNSVDCSSNNGSSSLQKKPISIVAKITGSTDTSRFGASPNLLSQCHS 737
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1514  QQQQQQQQTHQ-----QQQQQQQTHQ-----QQQQHQ 1528

QY      738  AHNHTTHQQQQQOOLSAQELGKSTRPKRGKYRVNDRSLVEAVKAVORGEMSVHRAGSYUG 797
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1529  QQH--HQQQQQQLTQQQ-----LQQQQ----- 1548

QY      798  VPHSLETKVYKVERHMRPKRE-----PKRPQRLVGLTGRANKLQLDK-LKAG 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1549  -----QQQQQQQQQLQQQQHQQQNQAANKLLIPKLTESM--MPSRPDQQLQHQKVL 1600

QY      845  PHGSGKSLNALKONNODAAAAAALAAAAAATNGKLPFLFEGPQALSFQPMFM--- 901
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 12:25:26 : Search time 50 Seconds
(without alignments)
2240.730 Million cell updates/sec

```

Title: US-10-016-768A-10
Perfect score: 6030
Sequence: 1 MHSSYEISLERVAECMGR.....ILHEKLAQIKAEVDQADQL 1165

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

```
Database :
1: pir_76:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4839	80.2	1221	2	T13683	probable transcript
2	311.5	5.2	2038	2	A43742	female sterile hom
3	302	5.0	2248	2	A35938	profilaggrin - hum
4	299.5	5.0	2175	1	S03170	homeotic protein C
5	294	4.9	1212	2	T13804	shs protein - fruili
6	283.5	4.7	1180	2	S69305	stirpe a/b protein
7	282.5	4.7	1768	2	T13349	parallel sister ch
8	280.5	4.7	1741	2	T13510	parallel sister ch
9	279	4.6	2150	2	S71629	sensory transduct
10	275	4.6	1436	2	S57338	forked protein 5.4
11	275	4.6	1449	2	S57237	forked protein 5.6
12	271	4.5	1365	2	S14871	suppressor two of
13	270.5	4.5	1217	2	S52714	sericinB - silkw
14	269	4.5	2271	2	P90073	hypothetical prote
15	261	4.4	1085	2	S65149	Gene phagequeak prote
16	259	4.3	1028	2	A56038	DNA-binding protea
17	257.5	4.3	1077	2	A44607	serine-rich protei
18	256.5	4.3	1733	2	B88921	hypothetical protea
19	249	4.1	1165	2	S62982	vacuolar protein V
20	249	4.1	1505	2	JC4851	hypoaldr-inducible
21	249	4.1	3180	2	T13828	CREB-binding prote
22	248.5	4.1	1063	2	D86731	hypothetical prote
23	244.5	4.1	883	2	S04122	puff 74E protein -
24	244	4.0	1655	2	T13998	Gene mastermind
25	242.5	4.0	1385	2	T24276	hypothetical prote
26	242.5	4.0	4776	2	E95206	cell wall surface
27	238.5	4.0	1403	2	S24548	homeotic protein F
28	237	3.9	2248	1	D42088	adenylate cyclase
29	236	3.9	873	2	B53225	ecdysone-induced p

30	234	3.9	663	2	S21912	BRcore-01-21 prote
31	234	3.9	2232	2	T34434	hypothetical prote
32	232.5	3.9	1457	2	T14577	protein kinase Yak
33	231.5	3.8	1398	2	T13741	hypothetical prote
34	231	3.8	1093	2	T18513	AE17 protein - hum
35	231	3.8	1354	2	B34598	ecdysone-induced p
36	231	3.8	2732	2	T20532	hypothetical prote
37	230	3.8	2518	2	A56922	transcription fact
38	229.5	3.8	864	2	A49070	ecdysone-inducible
39	229	3.8	1658	2	T13748	sex comb protein -
40	228.5	3.8	2738	2	E83320	protein F07A11.6 f
41	228	3.8	1060	2	S33641	homeotic protein z
42	228	3.8	1213	2	S16356	ovo protein - fruit
43	226.5	3.8	1072	2	T13232	dachhund protein
44	226.5	3.8	1104	2	S53130	probable membrane
45	226	3.7	2559	2	A56923	transcription fact

ALIGNMENTS

	RESULT	1
Tl3283	probable transcription factor E93 - fruit fly (<i>Drosophila melanogaster</i>)	
C:Species:	<i>Drosophila melanogaster</i>	
C:Date:	13-Aug-1999 #sequence_revision 13-Aug-1999 #ext_change 17-Nov-2000	
C:Accession:	Tl3283	
R:Bachrecke,	E.H.; Thummel C.S.	
Dev.	Biol. 171, 85-97, 1995	
A:Title:	The <i>Drosophila</i> E93 gene from the 9JF early puff displays stage- and tissue-spec	
A:Reference number:	Z17648; MUID:96018744; PMID:7556910	
A:Accession:	Tl3283	
A:Status:	preliminary; translated from GB/EMBL/DDbj	
A:Molecule type:	mRNA	
A:Residues:	1-1221 <BAE>	
A:Cross-references:	NID:U25686; NID:g886047; PID:g886048; PIDN:AAA83228.1	
A:Experimental source:	strain Canton S	
C:Genetics:		
A:Gene:	E93	
A:Cross-references:	FlyBase:FBgn0013948	
A:Map position:	3R	
C:Function:		
A:Description:	probably acts in a stage-specific regulatory hierarchy in the salivary gl	
Query Match	80.2%; Score 4839; DB 2; Length 1221;	
Best Local Similarity	84.0%; Pred. No. 4.8e-234;	
Matches	972; Conservative 25; Mismatches 90; Indels 70; Gaps 9;	
OY	1 MHISVYSIELSERVAEECGMRQKHVODKTCGHLNIEEQPIAIGSEDEPQYNHSSK	60
Db	1 MHISVYSIELSERVAEECGMRQKHVODKTCGHLNIEEQPIAIGSEDEPQYNHSSK	60
OY	61 EIOSGNPHCKTENHRLEOQHNGSQLLEEEDSENNOTSHDSRTPTPGATSTSPPEPI	120
Db	61 EIOSGNPHCKTENHRLEOQHNGSQLLEEEDSENNOTSHDSRTPTPGATSTSPPEPI	120
OY	121 DWRPSAKCNFCVNGRLTLTVAAOGKLVAESAATAITSSSTSNSHIHQDSDSNSASLPHNI	180
Db	121 DWRPSAKCNFCVNGRLTLTVAAOGKLVAESAATAITSSSTSNSHIHQDSDSNSASLPHNI	180
OY	181 SSSSSSSNNNSSGNRRARHIAAASARATPAATPANSLLEYLLKTORAPAKKTSWDSMAAOILA	240
Db	181 SSSSSSSNNNSSGNRRARHIAAASARATPAATPANSLLEYLLKTORAPAKKTSWDSMAAOILA	240
OY	241 QFSLLADFNLIINSLAQQQQQQQIASAVTPTTSEVSAAATSPALKOTPSPVDAPLDL	300
Db	241 QFSLLADFNLIINSLAQQQQQQQIASAVTPTTSEVSAAATSPALKOTPSPVDAPLDL	300
OY	301 SSRPSPNSSISGVKSVACATPTPSGRRAVSEEDLSRALQDVVANKLDARKSASQNHQ	360
Db	301 SSRPSPNSSISGVKSVACATPTPSGRRAVSEEDLSRALQDVVANKLDARKSASQNHQ	360
OY	361 RSLIDNMLFGMKHHDOEDHDGDELEDSDNDAAEAEVDSNASTPVVPAAEFARAQLRKLSHL	420

THIS PAGE BLANK (USPTO)

[illegible][illegible]

THIS PAGE BLANK (USPTO)

Dd 1601 PPQOS--PSDKLHENAALAAVASAQKLYOT-----FKANEQNLKNASS--WSSL 1648
Qy 902 -----POTNATN-----AYGLPNTIRTEAMRNPQ 925
Dd 1649 ASANSPOHTSSSSSSAKKPRMDSFOQFRNKAKRRDLKLEAEKEKQOKEAAEEQ 1708
Qy 926 ASNHGLMKSAODVENVYDGIIRKTLQASE---GNGSAGNGSNGNGHGHGHGH 982
Dd 1709 QRKHKKSSSSSLTSAVAQAIAAATAAATLGAALAAALASSASNPSSGSSGAG- 1767
Qy 983 ALLDOLLVKTRPRLPTTNRNNDYAAATCSSAGSEYKRSQPMGNVADIKERPLADSGGS 1042
Dd 1768 -----STSOQATTCGRDRD-----RDRERERERSGSGCG 1796
Qy 1043 SDEESHASHINNNSDLAHNKY-----KSGCGCGCGNGQOTNGCRSSRMTSRDSETDA 1097
1797 QSG-----NQNSSNSANSNGPSAGSGSGCGGCGGSPASAGPNS-----GGGCT 1842
Qy 1098 SPSKSGENGCOQNHKMDLNGSSSSSHIKCESEATGHS 1138
Dd 1843 ANSNGCGCGCGGCPAL--LNAGSNGSGVSGCAASNSNS 1881

RESULT 3
A:Accession: A35938
A:Species: Homo sapiens (man)
C:Date: 14-Dec-1990 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C:Accession: A35938
R:Gan, S.Q.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.
Biochemistry 29, 9432-9440, 1990
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.
A:Reference number: A35938; MUID:91064347; PMID:2248957
A:Accession: A35938
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-2248 <GAP>
A:Cross-references: GB:J02929
C:Genetics:
A:Gene: GDB:FLG
A:Cross-references: GDB:119912; OMIM:135940

A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat
*46-569/Region: filaggrin repeat
70-893/Region: filaggrin repeat
*1074-1397/Region: filaggrin repeat
F:1573-1896/Region: filaggrin repeat

Query Match 5.0%; Score 302; DB 2; Length 2248;
Best Local Similarity 19.0%; Pred. No. 3.2e-07;
Matches 249; Conservative 160; Mismatches 578; Indels 324; Gaps 47;

Qy 6 YEISLRYAEFCMG-----RQWKHYQDKLTGSHLIEEQPIAIGSED-----EP 52
Dd 572 YOVSTHEGSESAHGSABSTRRQGSRHDOARDSRHASQEGOPTIRGHFGSSRGCG 631
Qy 53 SQYHSSKEISQSNPNHCKT-----ENHRLLEQOHNSQLLEEDSENNQTSH----- 99
Dd 632 SHYEOSVDRSGHSGSHHTTSGCRSDASHGTSGRSASRQTRDEQSGDGRHSGSHHQ 691
Qy 100 -----DSSRTPTPCATSTPSPPRPIDMRPSAKNFCVNGRLTLVNAQGLVAESATA 153
Dd 692 EASTQADSSRSQYQ-----QG-----QSAGSR 714
Qy 154 TSSSTSNSHIHQODSDNSASLPHHISSSSSNNSSGNRAHTAASARATPAATPA 213
Dd 715 TSNQGSSESVQDDBEQSGESEDSERH---SGASRN-----HRGSAOEGSRDGRHPC 763
Qy 214 NSLELYKLLTORAAKMTGMDSWAAQLQPSLLADNLNLSLASQOQOQOQOQIASAVPT 273
Dd 764 SHDE-----DRAGHRQASADS-----SSQSGTRHTQT 789

Qy 274 TSEVSAALISPALKDTPSPVDABLDLSKSPSPNSISGDVKSVRACATPTPSCGRAYSE 333
Dd 790 SSRQAASSQEQASRAGDRHSGHQQSADSSRHSGI--GRQASAVADRGHGRSGQSA 848
Qy 334 EDLSRALDQVANKLDRKASQH---HQ--RSILNR-----LFKMHQHOEDQ-- 379
Dd 849 SDOEGHSESDSQSVSQRQAGSHQOQSHQOESTRSGRSGSGSFYOVSTHEGSESAH 908
Qy 380 -----HGDDLPSNDAAEAELVDSMASTPVYPAEFARQOJKLHSH-----LSEHN 424
Dd 909 GRTTSTGRGSGHNEQARQDSRHASQEOGDTIRAHPSRGRGQGSNHQEOSVDRSGHS 968
Qy 425 GSDLGEDVDRCSPKXGHRPACGNASANQAPASIPLDANVLTMLMLAIG---AMPKL 481
Dd 969 GSHHSTTSGCRSD--ASHQSGSGSHHQQS-----ADSSRH-----SGIGHQASSAV 1015
Qy 482 DETQTVDFITGLLVANSQGIIMEGLNLILSAS-----QENSG-----NAS 523
Dd 1016 RDSGHRGS--SGSQASDSEGHSESDSTQSVAGQACPHQOQSHQESTRGRSAGRSGRGS 1073
Qy 524 LLLQOQOHQOHQOHQOQOQOQVAAVYRHLRPKSETPETNSLDPNDASEDPIKIPSF 583
Dd 1074 FLYOVSTHEGSESAHGRARTSTRGQSGH---EQARDSRHSTSGEGDTRIIGHF-- 1126
Qy 584 KVSQPASSSSLSFGGLVGGHHRPLNNNSLSISNNSHNSHNRGNSRSPHASPMLAA 643
Dd 1127 ---GPRSS-----GPRHSHYEQSVN---STGSHSGHSHSTTSGCRDASHGTSGRSA 1173
Qy 644 A-----VAQGYASGNLSLTSSSSSIQKMAASNIQROINQESQOESLRN--GNVSDCSN 696
Dd 1174 SRETHNEQSGDGRHSGSRHQAESVWADSQAGQAGQEGSSGRTSRNQGSSSPSODSDS 1233
Qy 697 NGSSSLGYYKPRSTSVAKIIGCTPTSRFCAAPNLSQOHNHNLTHQOQOQOQSAQAL 756
Dd 1234 QGQSEDSERRRSGSASRNHRGASAREQSDGS-----RHPSGH--EDRAGHGSADSSR 1284
Qy 757 GKGTTPKRGKRYNRDRLDVLAVAVORGENSEVHRAGSYGVPSHTEYKVKERHLMRPR 816
Dd 1285 QSGTRHTQTSRR-----QAASSQEQARSRAGGRHSGHGO--QSAASSRH----- 1327
Qy 817 KREKRPQDVLGLTPANKLQDLKAKGPHGSKLSNLYKNQNNQAAAAAALAAAAA 876
Dd 1328 -----SGIRGQASTAVRDSGHRGSGQASDNEGHSESDSDTQSVAGORQAG 1374
Qy 877 TPNGKLRLPFAAGQALSFQNMF-----WPQTNATYAVGLDNRITTEAMRNPQA--SN 928
Dd 1375 SHHSHOESTRGQRETSGRSGFLYOVSTHEGSESSHGWGSTSTRGQSHNEQADDS 1434
Qy 929 HHGLMKSAODVENVYDGI-----IRKTLQASEGNGSAGNGSNGNGHGHGH- 978
Dd 1435 RHASQEGODTINHHPGHSAOSSRQSTRTTESSNGQAASSHNEQARSSAEGENHSGH 1494
Qy 979 -----GHCHALLDOLLVKKTRPFTNHRN-----NDYAATGSSASGESVYKSG 1021
Dd 1495 QSADSSRHAGIGHQQA-----SSAVRDSGHRGYRQSQATDSEGHSEDDTQSVSNOG 1546
Qy 1022 SPMGNVADIKREPLSADSGSGSDEEHSASH--INNNSDLAHNKKSGCGCGGG----- 1073
Dd 1547 Q-AGPHQAOHESARQSGSGSGSFLYOVSTHEGSESTHGSVPSGTGRQSGHHQDA 1605
Qy 1074 -----GNGQTNNGRSGSRMTSRDSETDASSFKSGENGCGOQNHKMD----- 1115
Dd 1606 QDSSRHASQEGOPTIRGHPESSR--GGRHSGHYEOSVDRSGHSGSHHSTTSGCRSDASH 1664
Qy 1116 -LNGSSSSSHIKCESEATG--HHSRQHHTTSLIHEKLAQIKAQVQDQADQ 1164
Dd 1665 GSSGSRASRQTRNEQSGDGSRRHSRSH-----HEASTQLESSRHQAGQ 1710

RESULT 4
S03170
homeotic protein cut - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster

THIS PAGE BLANK (USPTO)

[illegible]

```

Db      811 GFLPGSLAFQFAAAQVAAAGCGGRCGRHYRFADSLQLRPGASMSMAGRLGESLI PKGDPRMEAKL 870
QY      753 QEALGKGTGRPKRGKYRNYDRDLSLVEAVKAVORGEMSVHAGS-----YGVGHSTLEY- 805
Db      871 QEML-----RYMMDXYANQALDTL--HISRYVELLSVHNIGRLFAKYLIGLSQGTIVSEL 924
QY      806 -----KVKERHLMPRRKRE-----PKPQDYL-----VGLTGPAN 834
Db      925 LSKRPKPMDKLLEKGRDSYRKYHNAMACDNAVVLKSLIPKKSQGLPQYAKRGACGACGDD 984
QY      835 KLQDLKLKAGPHGSGSKLSNALK-----NONNQAAAAAATAAATAATPGLKLPFLF 886
Db      985 SMSEDRI--AHILSEASSLMKQSSVAAQHEQERSHSGCEDSHSNEDSKSPPOSTSPFF 1041
QY      867 EAGQALSLFGNNMFQTNATPAVGLDFNRITFQARNRPASNNHGLMKSAQODVENVYDG 946
Db      1042 KVENQLKQH-----HLNPEQAAAQOREFEREQOREBQOOR--LRHDDQDKMARLYOE 1092
QY      947 IIRKT-----LQASEGNSAAGNSGNSGNGCHGHCHGHALLDOLL--VKTP 994
Db      1093 LIATPRETAPRFLFSPSLTGCAGMPCGAANA-----FRMADENMHVFEREI 1143
QY      995 LPTNHRNDYAA-----TCSAGSEYKRSPPMGNTYA 1028
Db      1144 AKLQHQHQOQOQAAQAOAFQPFSSLMALQOQVNLGAQDLSLAAAKADIKLNG----- 1196
QY      1029 DIKERUSADGGSS--DEHSASHINNNDNLDLHNKKSGSG-----GGCGGNG 1076
Db      1197 --QRSSLSHSGSSSCSDGERDDAYPSS-----LHGRASEGGTAPRAPRPGPGTGGAG 1249
QY      1077 -----QTNQ-NGRSMRTSRHSDSETDASSFKSGENGGOQNHKMDLNGSSSSSHIKC 1128
Db      1250 APPTAAPPTGASSNSAAPPLSNLSILPPLASSQGEFPATASPLORM---ASITWSLI 1305
QY      1129 ESEAAATGHSFGCHHTTSLHKEKLAQKAEVQVQADQL 1165
Db      1306 TQPVTPPHSTPQRPPTKAV--LPRITQOQFPMFNNVL 1339

RESULT 5
T13804
shs protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13804
R:Treisman, J.E.; Lai, Z.C.; Rubin, G.M.
A:Title: Shortighted acts in the decapentaplegic pathway in Drosophila eye development
A:Release number: Z17767; MUID:96038094; PMID:755710
A:Accession: T13804
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1212 <TRE>
A:Cross-references: EMBL:L42512; NID:g833709; PID:g833710; PIDN:AAC41608.1
C:Genetics:
A:Gene: shs
A:Cross-references: FlyBase:FBgn0010460

Query Match 4.9%; Score 294; DB 2; Length 1212;
Best Local Similarity 20.2%; Pred. No. 3.8e-07;
Matches 235; Conservative 139; Mismatches 421; Indels 368; Gaps 44;

QY      46 AGSEDEPSQVYHSSKEIQSOSPNHCKTENHRLDAQHNGSO--LLEEDSENNOTSHDSR 103
Db      7 AASEBSGQHQQHQOQO-----OOHQHQOQOPLATTSVTAASTTVLANO 50
QY      104 TPTTGATSTPSPPE--PIDWRPSAKCNFCVNGRLTLTVAAGKLVESAAT--ATSSSTS 159
Db      51 SPTNSQASSPENSQALPLRLRQO-----SAAATVAAAATVAAITGCTS 96
QY      160 NSH1HQHSDNSASLPHHISSSSSNNSSGGR--ARHIAAASARATPAAATPANSLE 217

```

THIS PAGE BLANK (USPTO)

```

Db      97 000000HNRNSTSMFDRVNAKFKPSSNAGCGNNPVRNRNSMLTBSRGVTIGTGCNIRK 156
Oy      218 LYKLTORAKWTSNDMSMAOQLOAFSLADFNLIINSLSAQ000000-----OOIASAV 270
Db      157 LTKV-----SSLTSHHFAVVCYCPSPNITVONSNNAGSNALQDTTSESLRLNMKRVAAQA 211
Oy      211 TPTTSEVAAAISPALKDTPPSPVDAPLDLSK-----PSPNSSIGDVKS 316
Db      212 TPTT--VGRASSNSLSTSTSL-APKSSSSSGGSGNSTPQ000000LVSSNSSSSSNN 268
Oy      317 VRACATPPSGRAASEEDLSRALODVANKLDARKSASQHHQESILDNRLFKYKHQ 376
Db      269 FTKASSPNNNGARVSGAATSAATGTTAA-----AGSHHQPHHHHHHHHHHHHH 321
Oy      377 EQ----- 378
Oy      322 000000TSLSQHASLTVAGGSASAGGGGGGSGSSGTAAGCTTRKPKTTSFEITSVT 381
Oy      379 -----DHDGDELEDSDDAEAEV-DSNASTP-----VYPAEFARAO 413
Db      382 VGHPKLNAAGOTGDSADLDDESHTDNDRITDLENETPSMEDTFSKEVYYANNAJST 441
Oy      414 LRKLSLSEHNG-----SDLGED 431
Db      442 NARVPTSSQYGLVVVDPIAPSLGQTIQNVQVNSDNIINVGAVTPGCTKKKDDIKET 501
Oy      432 VDRGSP-----KMGK-----HPACGNASANOGAPASIPLDANVLHTLML 471
Db      502 QHRSRFRVVKIESTEPFKRGKMGKMDYLDHSSVNGGNGNNEKSTTSSEAAIAT-TDGC 560
Oy      472 AAGTGA--MKLDETQTVGDFIKGLLVANSQGINNEGILNLSA0ENSNGNASLLQO 528
Db      561 AAGVAGAGEAPAKHTTOS-----MLPPTOKLNE--NHLEANSIDANVNAVE000Q 609
Oy      529 00H00H00H-----H0000000HVA 549
Db      610 00000000QTVGNALTTLTPVALNRNRSSTVTRS PNAVTEFLSPNLLAQ00000000L- 668
Oy      550 AYRHLRPKSETPETNSSLDPNDAE-----DPILKIPSFKV 585
Db      669 -----FDSVNMNANAASSPNPAGDPNNMDYARTAAQMLHOTLOOLKOREADMDVP--- 716
Oy      586 SGPASSSSLSFGC-LVCGHHHPLNNNSLSISNNSHNSNH--RNGSNRSPHSASPMIL 641
Db      717 PGAGGYAVYQCGDSAVGA---ASNNSNAAAAATESQSLSTSYVQ000000PLSPALIT 772
Oy      642 A-----AAVAGGYSAGNSLTTSSSSSIQKMMASNIQ---RQINEQSGQESLRNGVVS 691
Db      773 POAAPTFAAVNAG--QSPNFQLE00000000Q0ATQIDGIVPQRPNPQ0000Q0TPQOSTA 830
Oy      692 DCSNNGSSSLGYKPSISVAKIIGTDTSRFGASPNILSQHHSAHHLTHQ000000L- 750
Db      831 Q0000000ATSAVTAAPPOQTSNTSNAAVTTGQGTMPRLS-----HMTSYEQ00PMLG 884
Oy      751 -SAOEALCKGTRPKRGKYRNYDRDLSVEAVYAVORGEMSVHRAGSYUG-----VPHSTL 803
Db      885 AAAAAAAGATA-----TSVAAPQAI-P--TIQLOSAPRTIADPQOLWMPPOOQ 931
Oy      804 EYKKEKRLMP-RKREKPORDLVGLTGPANKLOLDYLKAGPHGGSKLNALKNQNNQA 862
Db      932 000H0E0000P00000PLPANIASAGANNNSNLTLTNTV-VATGEATYVALTLTDEQA 990
Oy      863 AAAAAA-----AAAAATPGLKLPFEAGPOLSPQPMFMPQNTA--TNAYGDLFNIT 918
Db      991 TTAALAAAFATAAAAAATGATSAATAATQOIQLOOQ0PNAESETESASGSAVAID-NKIE 1049
Oy      919 EAMRNPASNHHGLMKSAQDWE 941
Db      1050 QAM--DLVKSH--LMIIVREEVE 1068

```

RESULT 6

569205

```

setipe a/b protein - fruit fly (Drosophila melanogaster)
C.Species: Drosophila melanogaster
C.Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C.Accession: S69205
R.Frommer, G.; Voithnueggen, G.; Pasca, G.; Jaekle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A.Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube gu.
A.Reference number: S69205; MUID:96203082; PMID:8612588
A.Accession: S69205
A.Status: preliminary; nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-1180 <PRO>
A.Cross-references: EMBL:U42403; NID:g1147788; PID:AA802355.1; PID:g1147789
C.Keywords: alternative splicing

Query Match      4.7%; Score 283.5; DB 2; Length 1180;
Best Local Similarity 20.6%; Pred. No. 1.2e-06;
Matches 236; Conservative 128; Mismatches 457; Indels 323; Gaps 41;

Oy      35 LNIHQPIAIAGSEDEPSQYNHSSKEISQSNPNHCKTENHRLSEQHNGSQLLEEDSEN 94
Db      3 LTRKQNEL-IVGSOQHSAITSSASSAGATSSPGLSQGNLVLA--GACATNTSTSN 58
Oy      95 NOTSHDSRTPTPGATSTP---SPPEPIDWRPSAKCNFCVNGRLTLTVNAQCKLVAESA 151
Db      59 SODSLNTPPTLLGLSRNPLQFAPRPAPPIAVPSPAAGPTFGYQTANAPRLHNSPA 118
Oy      152 TATSSSTNSHHQHDSDSNSASLPHHISSSSSSSNNSGGRANHIAASARATPAAAT 211
Db      119 TSEVTAPOPPV--ELDEYVDILQVQQLLDSSA-----AAAA 154
Oy      212 PANSLEYKLTORAKWTSNDMSMAOQLOAFSLADFNLIINSLSAQ00000000IASAVT 271
Db      155 AANNPTTEQSVQ00Q0QTVVQPHRHOQO-----Q0000000Q0VLAKPR 201
Oy      272 PTSEVSAASIPALKDTPSPSPVDAPLDLSKPS-NS 309
Db      202 PRINLOKATEYAAQIAQVSSSPGSRVLDVPSYVGNPYHNHTTPGEDVLALMFGSN 261
Oy      310 ISGDVKSVRACATPPSPSRRAVSEEDLSRALODVYANKLDA---RKSQHHQESILDN 366
Db      262 GTGV-----APGTPSAAM-----IMEGLETVAPTHNAFLLTETAAAHFNVLSTDT 309
Oy      367 RLFKKXHHQDQDHDGDELEDSDDAEAVDNSMASTPVYPAEFAQRLKLSLSEHNS 426
Db      310 CLFK-----TSTIAQSSGSPSTRTVYLSPAQFG-----SHGGA 343
Oy      427 DLGEDVDKSPKMGHNPACGNASANOGAPASIPLDANVLH--TLMLAAGIGAMPKLD 484
Db      344 STSN-----SLNSSSTAASSSTSSSLHNTTASAAAAAATAAATAAAN 386
Oy      485 QTVGDFIKGLLVANSQGINNEG-----LNLISAS0ENSNGNASLLQ00H00H 536
Db      387 NSV-----LRARQUTAPTCGSGPGHVAQPSATASGSRSSASHLSLNTSGQSPSS 440
Oy      537 QH000000Q0VAVYRHLRPS-ETPETNSSLDPRDAEDDLIKPSTKVGSPASSSSLS 595
Db      441 AVQVEAHKQLEA--LPQDLNTPVTTSSDIPSFQPTTVVEPRPTTGISIEDLSLE 496
Oy      596 P-----GGLVGGHHNPLNNNS--LSI-----SNNSHSSNHSRNSRSPHSASPMIL 641
Db      497 P0VTSVASPVLSSHSPKLEERSTPRALAIVEESSNNSCMNYPQHNNNNNNNTTSSST 556
Oy      642 AAAVAQGYISAGNSLTTSSSSSIQKMMASNIQRTINEQSGESLRNGVNSCCSSNNGSS 701
Db      557 TTTSKQ-----TTSESNTECVSGPGNHTQSHQO--QOOLQHNNTSSSNSGCHSH 605
Oy      702 SLGYKKPSISVAKIIGTDTSRFGASP-----NLISQO-----HNSAHNLTH--- 743
Db      606 000000-----0000HMSPOOYQ0HQIQLH0000FGYHHHHHHHHHHHNSQL 653
Oy      744 000000LSAQEALGKGTGRPKRGKYRNY-----DRDSLVEAVKAVQ 784

```

THIS PAGE BLANK (USPTO)

```

Db      654  OQIQQQQQQQQQQQQQHQQQRP LHHQQQQLQHQQQQQHQQHQQHQA1QHQQSQQAASKISYR  713
QY      785  GENSINRAGSYGVGVPHVSTLETKYKERHLMRPKRKEPRPQRPDLVLTG----PANKL-----  836
      714  G1FTT--TGAMNMAAAAAAAAAAAAAQQHQQQQQQHQQQLPSPLQGLVLAAPMSEPSLSGNSW  771
QY      837  ---QLDKLKAGP-----HGSKLSNALKNQNNQAAAAA-----AAAAA  871
Db      772  GLRSPDKTMQRPFLFSLPAHYATMQQQQQQQQQQQQQQQQQQAAAGAAPSPYDDGAAAAAA  831
QY      872  AAAAATPNGKLK--PLFEAGP-----QALSFQPMF  900
Db      832  AAGHAELLGLTMDCTPRLYLKQRPSTYAGASAGFGLDLHSHSEQQLQQQYVRSQRYQ  891
QY      901  WQPTNATNAYGLDQNRITTEAMRNQ-----ASHHGLKMSAQDMVENYVDGI  947
Db      892  WLDSPAD--YAQQQQQQQVQQVQQQQQQQQQTLLPQPTSSASSSNALGLVLPKQENYPD--  947
QY      948  IRTKLQASENGSAAAGSNGSNG-----HGHHGHGHALLPOLLVKK  992
Db      948  ---MQQS-SNGTGYSGSGSGSSAAAAAANAQAQLAETSPSTSKGHEILSCVYQS  1002
      993  T-PL  995
      |||
1003  TVPL  1006

```

```

RESULT 7
T13349
parallel sister chromatids protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13349
R:Gandhi, R.L.; Goldberg, M.L.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z17657
A:Accession: T13349
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1,1768 <AA>
A:Cross-references: EMBL:U40214; MID:g1100982; PID:g1100983; PIDN:AA91230.1
C:Genetics:
A:Gene: pasc
A:Cross-references: Flybase:FBgn0004655

Query Match      4.7%; Score 282.5; DB 2; Length 1768;
Best Local Similarity 19.4%; Pred. No. 2.3e-06;
Matches 238; Conservative 160; Mismatches 467; Indels 361; Gaps 45

Oy 48 SEDPSQYNNSSKKEISQSNPNHCKTENRLRLCOHNGSGLT-EEDESENNOTSHDSSRRPT 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 ASEQPEQPSASGKQKQKKPK-----EEKKLKPEAPPSSVILGRARVAVYREVDEDERYPT 258

Vy 107 PGATSTSPREPERIDMWRPASKNCVGNORLLTVNAQGLVLESAATP--SSYTSNHHNQ 165
Db 259 P-----TKDILIPKAGQPAEVAATATLAAASSSAFTSS 292

Oy 166 HDSDSNSSASLPHNISSSSSSNNNSGGRARHIAAASARATPAATPANSLELYLTQR 225
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 TFGCGPGESELPPTSAAPSASSTSSQLPS--ASGSASNPSPASRTPEHPVILRISKG 349

Oy 226 AAKKTSNDMSMAQLAQSLLDPNLINSLAQQQQQQQQQQLASAVTPTTSEVSAALISPA 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 TSLRVLSTDS--EEPPSSPAHQHQLNQLSTEEPARSCDETPAPATPKIT--VKPL 403

Oy 286 LKQDPSPVDAPRLDLSKSPNSISGVKSVRAACATPTPSGRARVSEEDLSRALQ---- 341
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 404 RPPRAADSVD-----GSSAAVAGSAGDSFEEKRSQSLPEPNEDEEEDEEDEEPEPI 458

Oy 342 -----DVNAKLDARKASQHNEDRSLLDNRLFGMKHNDDEQ 378
Db 459 NYCTVKISPDKPKERKLKLIKTDVIRAIATAKAAA--ESRSKSKSKSKH--KQL 513

```

```

QY      379 DHGDELEDSNDAAEAVDSNAPV-----VYPAEFARAQLKLSHSEHGSDL 428
Db      514 LAAGSGAAPASGATPAEINSEFTTPPHALISEANSQAQHPTPSHLHQHLHPQRSSAV 573
QY      429 -----GEDVDRCSPKKGRHPACGNASANOAGA-----                     456
Db      574 ISPTRSDHPDFDSSSVLGSISSKGNTSPOLLAAVOEDSCVIRSGSSVITSDLETSHQ 633
QY      457 ----SIPLDANVLHTMTL---ANGIGAMPKYDETQTGVDFIKGLLVANSGGIINEGLN 509
Db      634 SSLVAPPDSIERLESMMMTIDGAGTAASAASVPETP-----LOEDIILA 676
QY      510 LLSAEOENSNGNASILLQQOQHOOHHOOHHOOOQOQOQOQOQOQVAAYRRRLPKSETPETNSSLP 569
Db      677 VLRGVEPRNLGNTD-----PEPTEEDQQQPKRATRGRKA-----NNNVDV 720
QY      570 NDASEDPILKIPEFKVGSPASSSSLFCGLVGHHPHLNNNSLSISTSNSHSNSHRNG 629
Db      721 TPPEAT--TRTGRAKGADATTAAISP-----PTGKN-----FRGTGSKAKAQEV 765
QY      630 SNRSFHASPYLLAAVAQ-----CGYSAGNSULTSSSSSTOKMAMANIROI 676
Db      766 DMEVEDMTMTTPVANEBOLEOATPPRRGRHAARANNNNLASVNNINIINKIAAN-----L 820
QY      677 NEOSGOESLRNGVSDCSSNNGSSSLGYKKPSISVAKIIIGDTISRFGASPNLLSQOH 736
Db      821 SAKAEASRLAAGVA-----CGAARSYGRRKKNOQTVL-----QQE- 858
QY      737 SAHLTHOQOQOOLSAOEALCKTRPKRKGRANTDRDSLVAOVAORGENSVIRASSY 796
Db      859 -----PVPEBOETPDPAEE----DQPTPAK----- 878
QY      797 GVPHSTLEYVKKEBMLPRKREPKPOPLVGLTGPAANKLQDLKLKGP-----HGGSKL 851
Db      879 -IPTH-----HREHSPDHDPDDDELINNNSNNSSLIHDSSSSPPPREDKFKDKF 929
QY      852 SNALKNONNQAAAAAAAAAAAAATAPIGLK---LPFLFAOPALSFOPMFMFOTNATN 908
Db      930 KRTLLTLDTOGAANNAGAGAAAAAPRESSGEORGAVKLVISKKKSIFSRALVPSDOAEQ 989
QY      909 AYGLDLFNRITEAMRNPOASHHHGLMKSAQDMENVENVDGIIKTQLQASEGNCSSAANGSNG 968
Db      990 A-----TVAKRH-----LYKSSWD-----AALEANG--GGTNSDA 1017
QY      969 SNGNHGHGHGHGHALLDQLLYKTPLP---TNHRNNDYAATCCSSAGCEVYKSGSP-M 1024
Db      1018 SNASASGCVAGAKXDHHLHLAAGKSDGDFGDPSSNNNGSSSACSSEA--STLGDSPAL 1074
QY      1025 GNYADIKERELSDAGS-----GSSDEHSASHINNNSDLAHNKK 1065
Db      1075 G-----KISRLAGKKGVPATSTSSAPFLDLEPIAGELDERSMA-----GA 1116
QY      1066 SGGGGGG--GGNGQTNG--RSSHMTSRDSETDASSFKSGENGCOONHKMDLNGSS 1121
Db      1117 SAGGTGATTTGGGAGGGGAPVRVDRKTK--DYVPVRNVKTA-----HOIOEIGEYO 1167
QY      1122 SSASHKCESEAATGHSPCHTTSTL 1147
Db      1168 MDDVEYLIALQPNNPATRCLSL 1193

```

RESULT 8
t13610
parallel sister chromatids protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: t13610
R:Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*
A:Reference number: Z17668
A:Accession: t13610
A:Status: preliminary; translated from GE/EMBL/DBJ

THIS PAGE BLANK (USPTO)

A:Molecule type: DNA
A:Residues: 1-1741 <MUR>
A:Cross-references: EMBL:298269; NID:e1355202; PID:e1251076; PIDN:CAB10973.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0013432
A:Introns: 348/3; 1219/3; 1500/3; 1557/2; 1587/1; 1650/3
A:Note: EG:87BL.2

```
Query Match          4.7%  Score 280.5; DB 2; Length 1741;
Best Local Similarity 19.1%; Pred. No. 2.8e-06;
Matches 238; Conservative 157; Mismatches 449; Indels 403; Gaps 46;

48 SEDSEQVNHSSKELISQGNPNHCKTENRLEQNGNSQL--EEDSENNQTSHSSRTPT 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 ASEQPEQSSASGKQKQKPKP---EEKKLKPEAPPSRVLGARAAVYVREDEDERYPT 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 PCATSTPEPPEPIDWRPSAKNFCVNGRLITVNAQGLVAESAATAT--SSSTSMHIHQ 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
259 P-----TKDLIIPKAGQPEVAATATLAASSSAFTSS 292
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 HDSDNSSASLPHNIISSSSSSNNNSGNRARIHAAASARATPAATPANSLLEYLTOR 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 TFGSPGSPSLPPTTSASASASTSOLPS---ASGSANPPSASRTPEHPPIVLRISKG 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 AAKKTSMDMAAQLAQSFLADFLILNSLASGQQQQQQQOIAASAVTPTTSEVSAAIIPA 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 TSRLVSTDS---EEPPSSPAHQNLQNLQVTEEPARSGDETVPASTPKIT--VKPL 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 LKQTPSPSVDAPLDLSKRPNSNLSGQVKSVRACATTPSGRAVYSEEDLSRLAQ---- 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 RPTAADSDV---GSSAAVGAAGAGSFEERKSQLEPNEDEEEEEEDEEEPEI 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 -----DVANKLDARKASQHNHORSITLDRLEPKMHQDOEQ 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
459 NYCTVKISPKRPERKRLKIITVDIRNAIAKAAADA---ESRSKRSKSKKHG--KQL 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379 DHQDLEDSNDLAEVDSNASTP-----VYPAEFAARALRKLSHLSEHNSDL 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514 LMAAGCAAPAGCATPAEINSEFKTPSPHLALSEANSQQAQHTPSHLQHLHPQSGAV 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 -----GEDVDRGSPKMGHRPACGNASANOGARA----- 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
574 ISPTTRSDHDDSGSSVLSGSISSKGNSTPQLLAQVQEDSCVHSRGSSVITDLETSON 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
457 -----SIPLDANVLHTLTL---AAGIGAMPKLDDETQVDFIKGLLVANSGIMNEGILLN 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
634 SSVLAPPSDIESRLESMMMTIDGACTGLASAVPEP-----LQEDTLA 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
510 LLSASQENSNGNASLLQQQHQHQHQHQHQHQHQHVAAYRHR-----LPKS 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 VLKGEVPLNGNTD-----PEPTEBEDQQQPKRATRGGRKANNNDVTPPAT 725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
559 ET-----PETNSSLDP-----NDASDPLIKLPFVSG-PASSSSLS 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
726 ETRTRGAKGADATTAISPRTGKRNTGTRGSRKAEEVEVMEVDAMITVPANEEOLE 785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
596 PGGLVGGHHPHLNNNSLSISNNNSNHSNHSNRSPHSAPMLAAVAAGVSAGNS 655
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
786 QATL-----PPRRGRMAAARANNNN----- 805
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
656 LRTSSSSSIQKMMASNIQROIENGQGESLRNGNVSDCSNNGSSSLGYKKPSISYAKI 715
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
806 -LASVNNNNINKIAAN-----LSAKAEASRLAEGVA-----GGAARSTGRRKQOQVYOV 854
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 ICGTDTSRFGASPNLLSQOHHSAHHLTHQOQOOLSAQELGKGTGPKGKYRYNDRDSL 775
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
855 L-----QQE-----PVPEQETPDAAE-----EQPTPAK----- 878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
776 VEAVKAVQRGEMSVYRAGSYGVPHSTLEYKVERHMLRPKREPKRPDVLGTGPANK 835
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
879 -----IHTD-----HREHSPDHDPDPDPDLNNSNNS 908
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
836 LQDLKLKAGP-----HGGSKLSNALKNQNNQAAAAAATAATPNGLK--LPLFE 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
909 LQHDSSSSPPRPDPFKFKDKFRTLTDQGAANAAGACAAATAAPRESSSGEORGAVKLV 968
Qy          888 AGPALSQPMWMPQNTATNAYGLDFNRIRIEANRNPQASHHGIMKSAQOMVEVYGI 947
Db          969 SKKGSIFKSRALVPSDOAEQA-----TVAKRH-----LYKHSW----- 1003
Qy          948 IRKTLOASEGNSAAGNSGNSNGNGHGHGHGHGHALDQLLVKTPLPF---TNHRND 1004
Db          1004 -----ALLEANG--CGTNSDASNSASGVGAAGKDHHLHLACKSDDFDPSNSNNG 1056
Qy          1005 YAATCSSASGESVVRSGSP--NGNVAADIKREKLSADSG-----GSSD 1044
Db          1057 SSSACSSA---STLRGDSPALG-----KISRLAKOGVPATSTSSDAFDLLEPIAGELD 1108
Qy          1045 EEHSAHNNNNNSDLANNKXNSGGGGG--CGNQOTNG--RSSRMTSRDSEIDASSF 1100
Db          1109 LERSAA-----GASAGGTGATTGGGATGGGPIRVDRKTK--DYPPVVRNV 1153
Qy          1101 KSGENGGQNNHKMMDLNGSSSSSHIKCESEATGHSPPGHHTTSIL 1147
Db          1154 KTA-----HQIOEIGEYQEMDDVEYILDALQPHNPATRCLSAL 1193
```

```
RESULT 9
S71629
sensory transduction histidine kinase dhxa - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 24-Sep-1998
C:Accession: S71629
R:Wang, N.; Shaulsky, G.; Escalante, R.; Loomis, W. F.
EMBO J. 15, 3890-3898, 1996
A:Title: A two-component histidine kinase gene that functions in Dictyostelium developme
A:Reference number: S71629; MUID:96324397; PMID:8670894
A:Accession: S71629
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2150 <MAN>
A:Cross-references: EMBL:U42597
A:Experimental source: strain Ax4
C:Genetics:
A:Gene: dhxa
A:Map position: 6
C:Superfamily: response regulator homology
C:Keywords: autophosphorylation; phosphotransferase; two-component regul;
F:207/2142/Domain: response regulator homology <RKH>
F:2076/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match          4.6%  Score 279; DB 2; Length 2150;
Best Local Similarity 18.2%; Pred. No. 4.3e-06;
Matches 140; Conservative 144; Mismatches 323; Indels 164; Gaps 27;

495 LVANSGIMNEG-LNLNLSAQENSNGNASLLQQQHQHQHQHQHQHQHVAAYRHR 553
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13 IIGDTSVINTGDPNPLRTOQQLQQO-----QQQQQQQQQQQQQQQQQQQQQQQ 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
554 RLP-----KSETPNSSLDPNDASEDPILKIPSFYKSGRASSSSLSPGGLVGHHHP 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 HTPQOLYQKQOQOQSHSYGNHSFIHNVSPSYDINNNNNNNN-----NNNNNN 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
609 NNNLSISNNNSNHSNHSNRSPHSAPMLAAVAAG-----GYSAGNS 655
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 NNNNNNNNNNNNNNNNNNNNNNNNNNNNN--YYSPDIENSNTSKLSIEVVLNQFPINFLYNSNN 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
656 LRTSSSSSIQKMMAS--NIQROIENGQGESLRNGNVSDCSNNGSSSLGYKKPSISVA 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 NYLNNSSSLNHNINSVSLSNNNNNQTNQOPINNNNNNNNNNNNNNNNNNNNNNNNNNNN 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
714 KIIGTDTSRFGASPNLLSQOHHSAHHL---THQOQO--QQLSAQELGKGTGPKR---- 764
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 -----GNNNNNITDPT--KSKRSTYETNIGSHORRKSIGSLANSATHSFSKUKNPKLS 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 -----GKYRN-----YDR--DSLVEAVKAVQRGEMSVYRA 792
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

THIS PAGE BLANK (USPTO)


```

Db 292 SSTPSTVNTTCGAVNNNNNNNNNNSTGSLGAIIPMDRSDGGINITTEESTGNNSPRS 351
QY 793 -----GSYYGV-----HSTLEYKVEKHLMRPKRE--PKPOBDVL- TGPA 833
Db 352 NCGSNGCNGGIGLSPRLNLSLNGVAVSPRLNHLNNLNNNSNLPPLSPRLNFIHINVS 411
QY 834 NKLQDLKLAQPRGGSKLSALKKONQAAAA-----AAAAAAAAAATNGILKLPFEA 868
412 NLNNNNNNNNINPNNNNNNSN--NSNNNVSPRNNHNIISPRGSIISPRSNNGGTTI--- 465
QY 889 GPOLSFQPMFMPQTNATAYGLDFNRITFAMRNPOASHHGIMKSAQOMVENVDGII 948
466 SPRLNINNNIT---NNINNNNITLPPRLNPSRLNVPNTSPRLATSLNITLPIVSSL- 521
QY 949 RKTLOASEGNSAAGNSGNSNGHGHGHGHALLDQLLVKTPLPFTNHRNNDYAAT 1008
Db 522 -----TSSNNNNNSNNNTNPSINNNGR-NGHCIGTISEEILGNKPVVYNGNNNNNT 575
QY 1009 CSASGSESVKSGSPMGNTADIKERLS----- 1036
576 NNSTNNNNITNNNNNNNNNNNNNNNNVLTSPRKRTKGNHSTKNSLODETSSMNGDSDIS 635
1037 -ADSGS-----SDEHSASHINNNSDLAHNKNGKSGGGGGGNGQOTNGCRSSRMTS 1089
Db 636 GAGSGGSLRRNKDDNDENDGNSNTNNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 695
QY 1090 RDDSETDASSFKSGENGCGQNHKKMDLNGSSSSSHIKCESEATGHSFG 1140
696 NNNNNNNNNNNNNNNNNNNNNNNNN--NNNNNNNNNYHGATMMMSHNGSIG 744

RESULT 10
S57238
forked protein 5.4K - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 22-Jun-1999
C:Accession: S57238
R:Hoover, K.K.; Chien, A.J.; Corces, V.G.
Genetics 135, 507-526, 1993
A:Title: Effects of transposable elements on the expression of the forked gene of Drosophila
A:Reference number: S57236; MUID:94063487; PMID:8244011
A:Accession: S57238
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1436 <HMO>
A:Cross-references: EMBL:X69871
C:Genetics:
A:Gene: FlyBase:f
A:Cross-references: FlyBase:FBgn0000630
A:Annotations: 43/2; 104/3; 164/3; 564/3; 717/3; 825/1; 1001/2; 1121/3; 1204/3; 1274/2
perfectly; unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
J-83/Domain: ankyrin repeat homology <ANI>

Query Match 4.6% Score 275; DB 2; Length 1436;
Best Local Similarity 19.2%; Pred. No. 4.2e-06;
Matches 210; Conservative 146; Mismatches 369; Indels 368; Gaps 39;

QY 22 QMKHYODKLTCSH-----LNIEEQPIAIGSEDEPSQYNNHSSKEISQ 64
Db 136 RMGQSRKSLDKYKGRPIINDAENQVECLNVLVGHGTSVDYKSGSSQRRHKSQQQLHQ 195
QY 65 SNPNHCKTEHRLLEQHNQGLLEEDSENNQTSHDSSRTPT--PGATSTPSPPRPIDW 122
196 -----OHOOOQOQOQOQOHLSSCNSNSNSKTSRNTIKSKSSTLSSDVEPFYL 246
QY 123 RPSAKCNFCVNGRLTYNAQGLVAESAATATSSSTNSIHQNDSDNSASLPHHSS 182
Db 247 HPPILAGSGSGGGLCMGKMGKMGKMGKISKNSDA--LTSQOGRSSSEKLYNGSS 303
183 SSSSNNNSNGNRARHIAASARATPAATPANSLYKLTORAAKNTSMDSMAAQLAQF 242
Db 304 SLGGNSSSGG-----GVGGGGSGGALLRPDGLVYVPM--RNGGMYNTPSPVSGISGE 355

```

```

QY 243 SLL-----ADPNLINSLSAQOQOQOQOQIASAVTPPTSEVSAAIAPALKDTPSPVDAPL 298
Db 356 SFFLLDPDYHLIQSGAGSVGRGLQOQOQETAWQ----- 388
QY 299 DLSKPSPPSSISGDVKSVRACATPTPSGRAYSEEDLSRALQDVYANKLARKASQHN 358
Db 389 --ESRPPATVESGQRYDHQADVTAPA-----DGGSDDESIVISSASTRSSSSC 438
QY 359 EGRSLDNRLFKMKHHQDEQDHQDELEDSDDAEAEVDSNASTPVYPAEFARAOURLKS 418
439 AARVSI-----ATVAATSAP-----G 455
QY 419 HLSHNGSLDGEDYDRGSPKMGHRAAGNANOGAPASIPIDANVLTMTMLAAGIGAM 478
456 NTKNNN-----YKVTASP-----AIRAAGI--- 477
QY 479 PKLDETQTVGDFIKGLLVANSGGIMNEGILNLSAQENSNGNASLLQOQOQHQQH-00 537
478 -----TTT-----EDILYVRESRKQHQQOQOQHQQOQOQLOHQATQ 514
QY 538 HHQOQOQOQHVAAVRRHLRKSETPETNSSLDPNASEDPLIKIPFKVSGPASSSSLSFG 597
515 OQOQOQOQOQOQLOHKYVNGRSRSDGSH-----SRASASSSTRST 555
QY 598 GLV---CGHHPLNNNSLSISNNHNSHNRGNSRSPHASPWL----- 641
556 DILQYSNHH--LNRRNNNNNNNNNNSSNIAQSSSSNNNNNSLLNRRKSHITIGLHS 613
642 --AAVAGGYSAGNSLTS-----SSSIOKMAASNIORINOSGQESLRNGVSD 692
Db 614 SKYESCLKONYSKAVNVLKNQNLNGIKSDTYESVCPREDVAERTQTHKNSMIRNNLAD 673
QY 693 CSNNGSGSSSLGKKRSIS-VAKIIGCTDSRFGASPNLSQGHSAHLTHQOQOQOLS 751
674 ASSNNNTSGSIN--NISNIGMNGQOSSR-----NLKRVSSAPPMQNL- 715
QY 752 AQEALCKGRPKRGKRYNDRODSLEAVKA--VORGMSVHRAGSYVGVPHSTLEYKVK 808
716 --AVNNGPPP-----PRLPRLPAPVQOQNNNSVDQPN----- 748
Db 809 ERHLMPRKREPKRPQDVLGTGTPANKLOLDKLAGPHGSGSKLSNALKKONQAAAA-- 865
749 -----MTGPT--MHQKVPFGNGP-----TNANGCGGAAPP 780
QY 866 -AAAAAAAAAATPGLKLPYLEAGPQALSFOPNMFQPTNATNAYGLDFNRITFAMRN 923
Db 781 VPAPNPVATEAVDSDGLEV-----VEEPLRPSSELVGNHRTWSTISANKKAKLL-N 833
QY 924 PQAASHHGLMKSAQDMVENVDGIIRKTLQASGNSAA-----GNGSGSNGNGH 974
834 AGTSNOSTSIAASSDDS-QSRYYGGSVH-----AANSSAANGHFYGVSEGGKNGGNSVANGN 866
QY 975 GHGHHGHALLDQLLVKKTPLPFTNHRNNDYAATCSASGESVKSRSPPMGNTADIKRER 1034
887 GNGNGGG-----NYSHRNQPHYAT-----GQPHQYRPSLYG----- 919
QY 1035 LSADSGSSDDEH 1047
Db 920 -----CGSSAEDH 927

RESULT 11
S57237
forked protein 5.6K - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 22-Jun-1999
C:Accession: S57237
R:Hoover, K.K.; Chien, A.J.; Corces, V.G.
Genetics 135, 507-526, 1993
A:Title: Effects of transposable elements on the expression of the forked gene of Drosophila
A:Reference number: S57236; MUID:94063487; PMID:8244011
A:Accession: S57237

```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

[illegible]

RESULT 13
S52714
sericin1B - silkworm
C:Species: Bombyx mori (silkworm)
C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
Accession: S52714

R:Garel, A.A.; Delage, G.G.; Prudhomme, J.J.
submitted to the EMBL Data Library, March 1995
A:Description: Structure and organisation of the Bombyx mori sericin I gene and of the f
A:Reference number: S52714
A:Accession: S52714
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1217 <GAR>
A:Cross-references: EMBL:248802; NID:g755699; PID:g755700

Query Match	4.5%;	Score 270.5;	DB 2;	Length 1217;
Best Local Similarity	18.5%;	Pred. No. 5.7e-06;		

QY 5 SYEISLERRVAEECMGRQWKHYODKLTCSHLNIEEQPIAIGSEDEPQO-----YN 56
|| : || || : || || : ||
Db 158 STGVSAADRGSGASSRRROANYSDK-----DITAAISKDDSRADSSRRSNAYYN 205

5 / H-----SSKEISQSNPNHCKIEN--HRLDQDHNG-----SQLEEEBSENNQI----- 97
 Db 206 RDSGSESAGLSDRSASSSKNDNVFVYRTKDSIGGQAKSRSSHSQESDAYYNSSPDDGY 265

```

Db      : ||| : ||| : |
        : ||| : ||| : |
266 NAGTRDSSTNNKKASTIYADKQI-----RAANDRSSSQLOKSSAQIS 311

```

Db 312 SGPKGTVSSKDRQYSNDKRSKDAYVGRDGTVAAYSNNKDKSEKTSRQSNNTYVADQNSVRSD 371

OY 199 AAASAKATP-----AAATPANSLELYKLLTORAKKNTSMDSMMAQLAQFSLA 246

DQ 372 SPSASDQI3N61DNVLSZLNATVNLISGCGAGCGACGGGCTTNDNDGCGCCGTATGAGNCTC --- 120

DY 247 DFNILNSLASQQQQQQQQQIIASA VPTTSEVSAAAISPALKDPSPSVDAPLDLSSKPPS P 306
...|||::|||::|||

QY 307 NSSISGDKVSRACATPTPSGRAYSEEDL-----SRALOD----VWANKUDARKSASOH 357
||::| ||::| :::| : | ||||
AET YHWGGDGT ETCCTCFET EDLVNUGUVRTEDECDLTITKIVRTURFECAGSCGC AOG

```
QY      358 HEQRSILDNRLFKMKHHDDQEODHDGDELEDS-----NDDAEE-----VDSNASTP 403
        ||:::||::||::|
Db      500 SSSRSQSQESASVSSSSSSTLSEDSSEVDIDLGNLGWMNNNSDNKAKQRAAGCATKEASSS 559
```

Db 560 TQATTVCAGADDSADSYTWMNPRSSSSSSASASSSSGCGNVGGSQSSCGTSCGNARCH 619

620 LGTVSSTGTSNT--DSSKSAG-----SRTSGTSTGYGSSSHRGASVSTGSSSN 669

Dd 670 IDSSTKAGSSSTSGGTSTGYSSSHR-----GGSVSTGSSSNTD 709

Oy 571 DASEDPILKIPSFKVGSPASSSSSLPGCLVCGHHPLNNNSLSISNNSNHSNGRNGS 630

631 NPSPHSAPMLAAVACGG--YSAGNSLUTSSSSSIQKMASNIQRIINEQSGQESLRNG 688

QY 689 NYS-----DCSSNN-----GSSSLGYKKRPSISVAKIIGCTDTSRFCA SPULLSQQ 734
:
: : : : :
A12 SVSSTGSSSNMSTTKNAGSPRTSGTSTVYGSSSH-----RCGSVSSTGGSSSTDSST 864

QY 735 MHSAAHPLTHQOQQQQQLSAQEAALGKGTFRKRGK----YRNYDRDSLVEAAKVAQVQSEMSVH /90

Db 865 KNA-----GSSITGGSSITYGYSSDSRDSGVSSTGSSSNTDASTD 903

THIS PAGE BLANK (USPTO)

```
OY 791 RAGSYGVPHSTLEKYKVERHLMPKREPKPPQDVLGLTGTPANKLQDLKAG--PHGC 848
|||
D 904 LAGSSSTSGSSTGYSSDSR-----DGSVSTGSSSNTASTDADLACSSTGCG 950
|||
OY 849 SKLSNALKNQNNQAAAAAATAATPGLKPLFEAGPQALSFQPMFWPQTATN 908
|||
D 951 SSTYGVSSDSRDSGVSTGSSSNTDADTLTG-----SSTS 966
|||
OY 909 AYGDFRRTTEAKRNPPASHHGLMKSAQDMVENVYDGIIRKTLQASEGNGSAAAGNSN- 967
|||
D 987 G-----GSSSTGYSSDSRD-----GSVSTGSSSNT 1012
|||
OY 968 -----GSGNGHGHHGHGHALLDQLVKKTPLPFTNHRNDYATCGSSA-----SG 1014
|||
D 1013 DASTDLAGSSTSGSSSTGYSS-----SSNRDGSVATGSSSNTDASTTE 1056
|||
OY 1015 ESYVRSQSGSPMGNYADIKRE-RLSADSGSSDEEHASHINNNSDLAH-----NKNK 1065
|||
D 1057 ESTTSAGSSTEGYSSSSSHDGSVSTSDGSSSTGSGASSSASTAKSDAASSEDGFMMWRRK 1116
|||
OY 1066 SGGGGCGGGCGGTNGNRRSSRMTSRDSETPDASFSGGNGCGQNHKMDLNGSSSSSH 1125
|||
D 1117 S-----GSGHKSATVQSSSTDKTSDSASTDSTGASTTTSSGSS--STGSGSTSDA 1170
|||
OY 1126 IKCESEAAATGCHSPGHHTTSLHE 1149
|||
D 1171 SSTSSSVSRSHHS---GVNRLHK 1191
|||

RESULT 14
F90073
Hypothetical protein SA2447 (imported) - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: F90073
Mae, A.; Mizumachi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirumatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: F90073
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 112271 <KUR>
C/Cross-references: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetics:
A/Gene: SA2447

Query Match 4.5%; Score 269; DB 2; Length 2271;
Best Local Similarity 17.2%; Pred. No. 1.5e-05;
Matches 195; Conservative 200; Mismatches 506; Indels 236; Gaps 27;

OY 43 TAIAGS---EDEPSOVYHSSKEISQSNPNHCKTENHRLLEQOHNGSQLLEEDSENQTS 99
|||
D 1015 ISTSGSLASDSDSMVSSMSTSGSCT-----SESLSGSCTSDSDSKLSLS- 1064
|||

OY 100 DSSRPTPGATSTPSPPEPIDMRPSAKNFCVNGLLTVNAQGLVAESATATSSSTS 159
|||
D 1065 -----TSQSGSTSTST-----STFSASVRTSESGSTSGMSAGSDSMSTSTSDS 1110
|||

OY 160 NSHIIHQHDSNNSASLPHIISSSSSNNNSGGRAHIAAASAKATPAATPANSIELY 219
|||
D 1111 TS-----DSKASASTAS--SESIQSASTSTSGSVSTSLSTNSERTSTSVSTSLSTS 1164
|||

OY 220 KULTORAAKMTSDMSMAQAQFSLADPNLINSLAQQQQQQQQAASAVPTTSVSA 279
|||
D 1165 ESDSISSTSTSDSISALISSESTSLSESNSTSDSESOSASAFISESISESTSTSTS 1224
|||

OY 280 AAI-----SPALKOTPSVDAPLDLSKPSPNSSISGDKSVACATPTPSGRAYSE 333
|||
```

```
D 1225 ESVSSSTSESTSLSDSTSESGSTSTSLSNSTSGSASIS-----TSTISSESTSTFS 1276
|||
OY 334 EDLSRALQDVANKLARKASAS---QHNEQRSLIDNRLFMKJHNDQDHDGELED--- 387
|||
D 1277 ESVSTSLMSTSTSLSNSTSLSTSLSDSTSDSKDSLSTMSVSDSLSTSKDSISSTST 1336
|||
OY 388 -SNDDAEAEVDNSASTPVPVPAEFARAQLRKLSHSEHNGSDLEDYDRGS PKMGRPACG 446
|||
D 1337 LSGSTSESESDSTSS-----SESKSDSTMSISMSGSTSGSTST 1377
|||
OY 447 NASANQAPASIPLDANVLHTLMLAAGICAMPKLEDTQVDFIKGLVANSQGINNEG 506
|||
D 1378 STSLSDSTSTSLSLASAS-----MNSQ 1398
|||

OY 507 LNLVLSAQENSGNMSLLLQQQQHQHQHQHQHQHQHQQAAYRHR-LPKETPEINS 565
|||
D 1399 GVDNSASQASNSTST-----STESDSQSTSTYTSTSTSTSTSTSTST 1443
|||

OY 566 SLDPNDASEDPILKIPSFVKVGPASSSSLPGLVGGHHHPLNNNSLSISNNSHNS 625
|||
D 1444 SLSDS-----TSISKSTSGSSTSTASLSGSESESDSQSISTASSESTSEAST 1493
|||

OY 626 HRNGSNRSPHSASPMILAAVAQGYGAGNSLTTSSSSSIQKMASNIQROI NEOGQESL 685
|||
D 1494 SLSDSTSTSNSGSASTSTSLSN--SASASESDSSSTSLSDSTASMSQSESDSQSTAS 1550
|||

OY 686 RQNVNDDCSSNNNGSSSLGKKRPSIAKIIIGDTPRFGASNNLLSQOHSAHLTHQ 745
|||
D 1551 LSLDSTSTSNR-----MSTIASLSTSVSTSESGSTSESTSESTSTSLSDSQ 1599
|||

OY 746 Q-QQQLSAQELGKTRPKRKYRNYDRDSLVEAVAVORGMSVHRACGYGVPHSTLE 804
|||
D 1600 STSRSTASAGSASTST-----STSDSRSTASASTSMR 1632
|||

OY 805 YKVKERHLMPKREPKPPQDVLGLTPANKLQDLKACPGHGSKLNALKNQNNQAAA 864
|||
D 1633 TSTSDSQMSLSTSTSTMSDSTSLSDPSVSDSTSDSTASSTGSMMSVSLSDSTSTST 1692
|||

OY 865 AAAAAAAAAAATPGLKPLFEAGPQALSFQPMFWPQTATNAYDLNRTTEAMRN 924
|||
D 1693 ASEVMASISDSQSMSESVNDSSESVSESNESDSK--SMGSGTSVSDSGSLVSTSLRS 1750
|||

OY 925 QASNHGHLMKSAQDMVENV-----YDGIIRKTLQASEGNGSAAAGNSGNGHGHC 975
|||
D 1751 ESVSESSSLSGSOSMSDVSSTSDSSLSVSTSLRSSESVSESOSSLSKSTSGS----- 1804
|||

OY 976 HGHGHGHALLDQLLVKTPLPFTNHRNDYATCGSSAG-ESYVRSQSGSPMGNYADIKR 1034
|||
D 1805 -----TSTSGSLSTSTSLSGSESVSESTSLSDSIS----- 1836
|||

OY 1035 LSAADSGSSDEEHASHIN-----NNNSDLAHNKKKSGCGGCGGCGGTNGNRRSSRM 1087
|||
D 1837 -WSDSTSTSDSDSLSGSISLSGSTSLSTSDSLSDSKLSLSSQMSGSE-STSTSVSDSQS 1894
|||

OY 1088 TSRDSETPDASFSGGNGCGQNHKMDLNGCSSSSSHIKCESEAAATGCHHSPGHHTT 1144
|||
D 1895 SGTSMNQFDMSTISASESDSMSTSDSNSTSGNSTSTSLSTSDSMGSGSVSTSTST 1950
|||

RESULT 15
S66149
gene p1psqueak protein A long form - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 23-Sep-2002
C/Accession: S66149; S66150; T45461
R/Weber, U.; Siegel, V.; Mlodzik, M.
EMBO J. 14, 6247-6257, 1995
A>Title: p1psqueak encodes a novel nuclear protein regulated downstream of seven-up for ci
A/Reference number: S66148; MUID:36134923; PMID:8557044
A/Accession: S66149
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1085 <WEB>
```

THIS PAGE BLANK (USPTO)

[illegible]

Search completed: October 29, 2003, 12:30:34
Job time : 140 secs

THIS PAGE BLANK (USPTO)